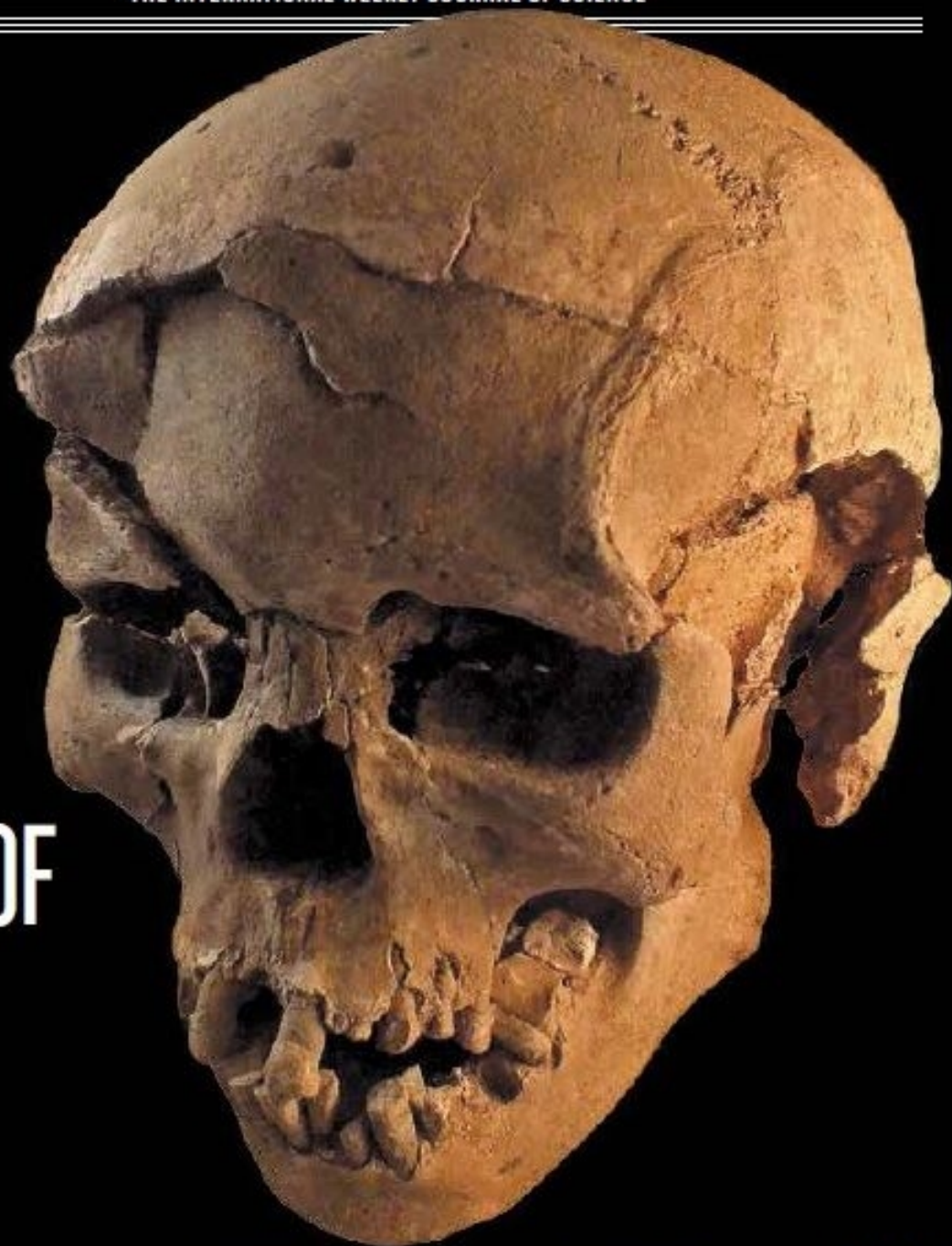


nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



Human
conflict
10,000 years
ago on the
shores of Lake
Turkana

PAGE 394

ECHOES OF ANCIENT WAR

GENETICS

GOING WILD IN HAWAII

*The tangled inheritance of
Kauai's feral chickens*

PAGE 270

BIOMIMICRY

LIFE LESSONS

*Engineers and chemists
should work with biologists*

PAGE 277

EXOTIC ATOMS

ANTIMATTER PUT TO THE TEST

*A new limit to the 'zero'
charge of antihydrogen*

PAGES 294 & 373

NATURE.COM/NATURE

21 January 2016 £10

Vol. 529, No. 7586



THIS WEEK

EDITORIALS

NEW STANDARDS Funders and society target sloppy scientists **p.256**

WORLD VIEW A first-hand account of sexual harassment **p.257**

GIRAFFE Fossil shows short neck but gigantic body **p.259**



Harassment victims deserve better

Sexual harassment is rife in science. Universities must stop trying to save face: they must discipline perpetrators and support victims.

How many senior scientists — usually men and usually with significant power over the careers of those in their labs — have been sanctioned and disciplined by their universities for sexual harassment? Nobody knows, especially not young researchers who eagerly apply for their first jobs, spend long hours on fieldwork and feel under pressure to socialize and make contacts after hours and at academic conferences. How many times have colleagues turned a blind eye to inappropriate comments and actions, and made excuses for people who should know better — and who are morally, legally and contractually obliged to behave better? How many young scientists have left positions, or left science completely, because of such behaviour, or because it is seemingly not taken seriously?

We don't know the answers to those questions. But one thing we do know is that sexual harassment is a serious problem in science. And we know that young female scientists are speaking up about it. We know this not because universities are being transparent about such complaints and how they are dealt with, but because, dissatisfied with the official responses, victims, journalists and others are bringing the facts about these complaints to light.

On page 257 of this issue, for example, *Nature* publishes the testimony of a female researcher who was persistently harassed by a senior male colleague. His university investigated and upheld her complaint. But it told her to keep the matter confidential, and although it promised action against him, allowed the offender to stay in his post. *Nature* knows who he is, but in this case, the female researcher did not want to name him for fear of reprisals.

Apologists for sexual harassment will tell you that it 'is rarely a black and white issue' and that inappropriate behaviour often 'falls into a grey area'. Read this woman's story: having an influential male colleague 30 years your senior ask to stay at your house for a work trip, request kisses and then enquire whether his night-time masturbation kept you awake is 100% wrong.

Nature and others have encouraged scientists to stand up to such behaviour. But it is clear that the system is weighted towards protecting powerful faculty members at the expense of students and young researchers. Although institutions proclaim that they have zero tolerance for abuse of the policies that they claim to enforce, too often their primary concern seems to be secrecy and reputation management.

A string of cases in the US astronomy community demonstrates this. In each, a university investigated sexual-harassment claims against a faculty member, found the claims substantiated and attempted to bury that fact from public view.

The latest disclosures, made public last week, revealed that astrophysicist Christian Ott of the California Institute of Technology (Caltech) in Pasadena was suspended without pay last year for harassing two female graduate students. And Congresswoman Jackie Speier (Democrat, California) took the extraordinary step of decrying sexual harassment in science on the floor of the US House of Representatives.

She entered into the *Congressional Record* a 2005 University of Arizona finding of harassment regarding Timothy Slater, an astronomy educator who later moved to the University of Wyoming in Laramie.

These incidents follow probably the most high-profile recent case, which saw exoplanet hunter Geoffrey Marcy leave the University of California, Berkeley, late last year, but only after complaints and a university finding against him were revealed by news media.

"Any principal investigator who thinks, 'It cannot happen at my university,' is wrong."

Scientists accused of such behaviour have the right to have their identities protected unless and until the claims are proved. But once an investigation has been completed and signed off, it is incumbent on those in power to be sure that they act on it. Disciplinary action is a good first step; ensuring that the victims have a path forward in science is another.

There have been baby steps in the right direction, such as the effort by US President Barack Obama's administration to make it clearer to students what their rights are and how they can go about reporting a sexual violation. And Speier is working to force the US Department of Education to make sure that when a person found to have violated the law changes institutions, all institutions involved are aware of the situation.

Any principal investigator who thinks, "It cannot happen at my university," is wrong. These are not one-off cases. They are examples of a systemic underlying rot that is driving many young researchers out of science for good. ■

Blue future

Coastal wetlands can have a crucial role in the fight against climate change.

Over the past decade, scientists and policymakers have joined efforts to create a science-based framework under the auspices of the United Nations to protect our remaining tropical forests. These carbon-rich ecosystems help to moderate the climate and serve as a treasure trove of biodiversity and a resource for local and indigenous peoples. Governments across the tropics have begun to incorporate forest conservation into their climate and development plans. Now it is time to do the same with coastal wetlands.

Roughly half of the world's carbon emissions are absorbed by the oceans, and the UN estimates that at least half of that carbon sequestration takes place in 'blue-carbon' wetlands. Often occupied by seagrass and mangroves, these saltwater ecosystems promote healthy fisheries and sequester carbon in their soils. Mangroves also stave off

erosion and serve as the first line of defence against powerful storms as well as saltwater intrusion into local groundwater resources. The world has lost more than one-third of its mangroves over the past several decades, and more succumb each year to shrimp farms, rice paddies and palm plantations, as well as to tourism and real-estate development. There's money to be made, but it's the environment that pays.

Nascent efforts are under way to halt this degradation, and a few pioneering projects have already shown success. Senegal is home to the world's largest mangrove restoration project, which began in 2008. Villagers have planted around 79 million mangrove trees across more than 7,900 hectares. The project has been registered and certified under the Kyoto Protocol's Clean Development Mechanism (CDM), and is benefiting from the sale of carbon credits.

In 2010, the United Nations Environment Programme launched the Blue Carbon Initiative, which seeks to reverse current trends and increase the area of coastal wetlands under effective management by 2025. The global climate agreement signed in Paris last December opens the door to advance such efforts, for example by enabling carbon trading and a programme similar to the CDM that allows countries and companies to pay to reduce emissions or build carbon stocks in projects such as the one in Senegal. It will be up to governments to incorporate coastal management into their climate plans, and to begin creating what some have called the 'blue-green economy'.

The available evidence justifies the pursuit of these efforts. Mangrove ecosystems alone could store as much as 20 billion tonnes of carbon — equivalent to more than 2 years of global carbon emissions — in their soils, much of which would be released into the atmosphere if the trees were destroyed. A 2012 study suggested that mangrove conservation could be effective at a cost of just US\$4–10 per tonne of carbon dioxide, which is within the current range of prices on the

European carbon trading system (J. Siikamäki *et al. Proc. Natl Acad. Sci. USA* **109**, 14369–14374; 2012).

In some cases, mangrove protection and restoration could even benefit from the existing forest-carbon-trading framework, which enables developed countries to invest in efforts to reduce deforestation in the developing world. But more science is needed, both to document the extent and causes of the problem and to provide the data that will be

“The planet that humanity calls home already knows how to sequester carbon.”

needed if countries are to incorporate coastal wetlands into their carbon inventories and climate planning. We know too little about what happens to the carbon locked up in plants and soils when they are converted for other uses.

Just as occurred with remedying tropical deforestation, science and policy can move forward in parallel. As countries establish coastal management policies, they will help to drive the development of both science and policies. One opportunity is in the Dominican Republic, which has devised a comprehensive plan to reduce emissions by conserving and restoring mangrove forests. That project is registered with the UN, and it incorporates scientific objectives, including quantification of the carbon sequestration and storage capacity of these ecosystems. This will inform the policy framework and provide the scientific basis for any economic returns that the initiative may reap years and decades into the future.

Meeting the objectives of the Paris agreement — to contain global warming over the course of the twenty-first century — will require urgent action on all fronts. Countries must work to reduce industrial carbon emissions, but ensuring that natural ecosystems continue to function is equally vital — and relatively simple. The planet that humanity calls home already knows how to sequester carbon. Let's make our forests and coastal wetlands work for us. ■

Repetitive flaws

Strict guidelines to improve the reproducibility of experiments are a welcome move.

From next week, scientists who submit grant applications to the US National Institutes of Health (NIH) will be asked to take a little more care. As part of an increasing drive to boost the reliability of research, the NIH will require applicants to explain the scientific premise behind their proposals and defend the quality of their experimental designs. They must also account for biological variables (for example, by including both male and female mice in planned studies) and describe how they will authenticate experimental materials such as cell lines and antibodies.

These demands are timely, sensible and, if researchers have been following the advice of their scientific societies, will sound familiar. Over the past year, a string of organizations have published their own statements and guidelines to boost the reproducibility of research.

Collectively, the message is: show your work, and don't fool yourself with unreliable reagents or shoehorned data. Updated guidelines from the Federation of American Societies for Experimental Biology, for example, call for standard ways to cite antibodies and animal-care practices. The Society for Neuroscience has asked for random sampling of everything from subjects to cell parts whenever an entire population is not studied. The American Psychological Association has called for infrastructure and policies to promote data sharing. The Biophysical Society has detailed how to make experimental data widely accessible. And the American Society for Cell Biology has called for subdisciplines to create community standards for assays. More guidelines are in the works, and funders and journals have weighed in too.

These will help to create studies that are more sound. Other biomedical funders should follow the lead of the NIH and introduce similarly tangible requirements.

The NIH has stated that communication and awareness are crucial to address the lack of reproducibility in research. But unreliable work has many causes — and aspects of today's scientific landscape can thwart quality research. Competition for funding and faculty positions, and for the publications necessary to secure them, encourages uncritical acceptance of results. All too often, it is better to be first but wrong than scooped and right. Journals, including this one, have gone some way to acknowledge and take up their responsibility for this (see go.nature.com/huhbyr).

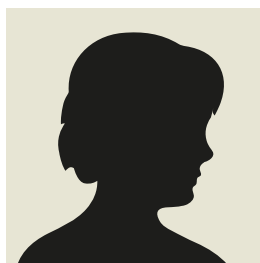
At the same time, experiments have become outsourced. Kits and reagents bought from commercial vendors allow scientists to do more research in a fraction of the time needed for 'home-brew' experiments with reagents created in the lab. These resources are invaluable but leave scientists less able to anticipate and identify artefacts.

Guidelines can cut down on mistakes from rushed publications and help to disseminate knowledge that would otherwise have to be gained by experience. But the real power of such recommendations lies less in their specific contents and more in the values that produced them.

Civil society depends on people acting according to societal norms, even when not doing so is unlikely to bring punishment. Citizens' motivation is not to gain material resources but to maintain their integrity. Similarly, the scientific community depends on researchers who adhere to values that are embodied in guidelines and recommendations. Doing quality research offers intrinsic rewards.

Guidelines work best when they build a culture that makes proper behaviour second nature. They can help to make researchers aspire to the values that produced them. They are valuable not only because scientists follow them but also because they can inspire researchers to uphold their identity and integrity. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunq



Sexual harassment must not be kept under wraps

A female scientist who was harassed by a senior male colleague feels let down by the system that is supposed to protect her.

As cases of sexual harassment in science receive more attention, female researchers are being urged to report such behaviour. We are told by our employers that they do not tolerate sexual harassment. My experience shows that this is not the case, and that universities need to update how they deal with such complaints. I was sexually harassed by a male colleague, my former postdoctoral supervisor, who is 30 years my senior. Several years after I complained, I still suffer the professional and personal consequences of having stood up to him.

The harassment began when I was working at a university in a different country from my former supervisor, but continuing to collaborate with him. He was planning a work trip and asked if he could stay at my home to save money and so that we could work on publications together. The harassment during his visit was daily and persistent. He made suggestive and lewd comments, such as asking me one morning whether his masturbating in the next room had kept me awake. He would try to kiss me. Each time I would ask myself: "How do I say no without damaging my career?" On the final occasion, I told him repeatedly that I found the aggressive way that he was talking to me intimidating. He then questioned why I was being so prickly around him and why I was shaking. My response triggered more verbal abuse, so I locked myself in a room and contacted a friend for advice. The following morning, he packed his bags, went through my possessions and left without a word.

I was reluctant to make an official complaint in case it made things worse. I decided to restrict communication to e-mail, rather than using Skype, which was a change in working practice that I had to justify to colleagues whom we worked with in his department. A few weeks later, I learned from one of these collaborators that, without permission, the harasser was taking my data, PowerPoint presentations and conclusions and using them to write a grant proposal. The colleague suggested that my name be included on the grant.

When my boss questioned why, less than one working day away from grant submission, our university had not yet received the paperwork necessary to approve my involvement, I told him about the sexual harassment. The harasser subsequently informed me that it was simpler to submit the grant without me. At this point, my boss advised me not to attend an upcoming field expedition, because my harasser was the lead scientist. My boss told the organizers that I was withdrawing for personal reasons.

A few months later, my harasser published my data sets in a journal article without my permission and without my name on the author list. My boss advised me to submit a formal complaint.

When I did — to his university — my harasser responded with dozens of pages of denials and counter-complaints, to which I was expected to respond. He belittled me, demanded access to my data sets, misrepresented evidence and argued for restrictions that would significantly detriment my career. Because of these issues and the confidential nature of the accusation, I found it nearly impossible to publish during the lengthy complaint process (which took much

longer than laid out in the university's own grievance procedures). I felt I had to excuse myself from international conferences, because I knew that he would be there. His career continued unaffected.

After almost a year and a half, his university told me that it had found in my favour. It said that he was guilty of both research misconduct and inappropriate behaviour, including sexual harassment. It did not fire him and stressed that I should keep the verdict confidential.

Although many universities claim zero tolerance of sexual harassment, this phrase is empty until the perpetrator's actions become enough to warrant dismissal. The confidentiality (secrecy) around verdicts leaves other women unaware that there is a confirmed harasser working at or visiting their department, teaching their lectures, leading their field trips or conversing with them at conferences. The secrecy leaves victims, like myself, unable to explain to collaborators, colleagues, funding bodies and potential employers why their CVs look so lean during that time.

Recently, my boss and I contacted the organizers of a scientific conference to enquire whether measures could be taken to ban my harasser from the event so that I would feel safe in attending. We were told that, because the outcomes had not been disclosed publicly, they were unable to act.

Until victims feel able to speak up freely, and details of such situations are shared, it is not the guilty parties who are punished, but young scientists. I was willing to put my name to this account, but I was reluctant to identify the offender. *Nature's* lawyers advised me, rightly, that this would be unfair on other male colleagues, who might find themselves subject to speculation. Hence, to prevent other men from being falsely accused, I chose to remain anonymous. ■

*The author is a researcher at a major university.
e-mail: nature@nature.com*

UNTIL
VICTIMS
FEEL ABLE TO
SPEAK UP
FREELY,
IT IS NOT THE
GUILTY PARTIES
WHO ARE
PUNISHED,
BUT YOUNG
SCIENTISTS.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/stodoy

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

PHOTONICS

Practical source of single photons

Single-photon sources are essential for quantum computers that encode information in light. Now, physicists have produced photons with high quality and at high rates.

Chao-Yang Lu and Jian-Wei Pan at the University of Science and Technology of China in Shanghai and their colleagues used a laser to excite an artificial atom in a semiconductor crystal, known as a quantum dot. By using finely tuned laser pulses, they produced individual photons with near-perfect uniformity. They crafted the crystal into a tiny pillar-shaped cavity to maximize the number of photons that escaped.

The device emitted 3.7 million high-quality photons per second — a rate that makes it good enough for practical applications, say the authors.

Another group, led by Pascale Senellart at the University of Paris-Saclay and her colleagues, has achieved comparable success using a similar technique.

Phys. Rev. Lett. 116, 020401 (2016); preprint on arXiv <http://arxiv.org/abs/1510.06499> (2015)

CLIMATE CHANGE

Oceans take a lot of heat

Of the heat taken up by the world's oceans since 1865, nearly half has been absorbed in just the past two decades.

Peter Gleckler at the Lawrence Livermore National Laboratory in California and his colleagues examined data on ocean temperatures from ship-based



ESTHER HERRMANN

ANIMAL BEHAVIOUR

Chimp friendships based on trust

Trust does not seem to be unique to humans and may have a long evolutionary history, according to a study in chimpanzees.

Jan Engelmann and Esther Herrmann at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, identified captive chimps (*Pan troglodytes*) that were friends with each other — individuals that frequently groomed and spent time together. The animals were then paired up and each chimp was presented with a choice of two ropes that it could

pull. The 'no-trust' rope gave the chimp a small food reward and left the partner empty-handed. The 'trust' rope gave the partner a reward, part of which could be sent back to the other chimp.

Chimps were more likely to choose the trust rope when paired with friends than with non-friends, showing that trust in close relationships is not unique to humans. The characteristic might have evolved to stabilize social bonds in primates, the authors say.

Curr. Biol. <http://doi.org/bbvs> (2016)

IMMUNOLOGY

Immune cell goes awry with age

Inflammation increases as the body ages, and one cause could be changes in a type of white blood cell.

Increased inflammation throughout the body is linked to age-related diseases from dementia to heart disease, and growing levels of an inflammatory protein called TNF are a known culprit. Dawn Bowdish at McMaster University in Hamilton,

Canada, and her colleagues studied the effects of TNF in aged mice. They found that the protein boosted the production of immature and dysfunctional monocytes, a type of immune cell, when the mice were infected with the bacterium *Streptococcus pneumoniae*. These immature monocytes drove further inflammation, and the infected animals were less able to rid themselves of the bacteria than were young mice. Blocking TNF in old mice restored this ability.

Antibacterial treatments

may need to be tailored to the age of the infected person, the authors say.

PLoS Pathog. 12, e1005368 (2016)

ORGANIC CHEMISTRY

Spring-loaded bond adds variety

Chemists have designed a simple way to attach small, strained-ring-shaped chemical groups to potential drug compounds — previously a difficult process.

The structures can change a drug's properties, for example to improve how it is absorbed in the body. Phil Baran at the Scripps Research Institute in La Jolla, California, and his colleagues attached a few such groups — including propellane, which looks like a propeller — to drug molecules by first breaking a high-energy carbon–carbon bond in the ring structure. The team used the energy released from this 'spring-loaded' bond to add another nitrogen-containing molecular group, creating a molecule called an amine. Using their method in a high-throughput synthesis, the team added strained molecules to a range of pharmaceutical compounds.

The team says that scientists in the drug industry have started using the method. **Science** 351, 241–246 (2016)

NEUROSCIENCE

Chemical revives neuron function

Treating regenerated nerve fibres with a chemical can restore vision in mice.

After injury, adult neurons regrow their long, signal-conducting axons to form connections called synapses with other neurons. But the regrown cells cannot conduct electrical signals properly. Michela Fagioli and Zhigang He of Children's Hospital Boston in Massachusetts and their team studied mice with injured optic nerves. Deleting two genes that

regulate neuronal growth, or boosting the production of two protein-growth factors, regenerated retinal axons that formed synapses in the brain, but did not improve vision. Treating the neurons with a compound that improves their conductivity helped them to work properly. This restored vision in six of the eight experimental animals.

The results could help to improve nerve-cell function after an injury, the authors say. **Cell** 164, 219–232 (2016)

PALAEOLOGY

Extinct giraffe was a huge beast

An extinct relative of the giraffe that lived more than 1 million years ago weighed up to 1,800 kilograms, making it one of the largest cloven-hoofed mammals ever to have existed.

Remains of the relatively short-necked giraffid *Sivatherium giganteum* were first uncovered in the 1800s, but until now, nobody had accurately estimated its body mass. Christopher Basu at the Royal Veterinary College in Hatfield, UK, and his colleagues created a 3D model of this giraffid's skeleton (pictured), which indicated that it weighed between 850 and 1,800 kilograms, with males carrying heavy horns.

The team concludes that the creature was not as large as modern African elephants, as was previously suggested,

but it may still have been the largest-ever four-stomached mammal, or ruminant.

Biol. Lett. 12, 20150940 (2016)



SOCIAL SELECTION

Popular topics on social media

Insider view of faculty search

Scientists hunting for academic jobs got a rare glimpse into the mysterious tenure-track hiring process. A blog post written by computational genomicist Sean Eddy at Harvard University in Cambridge, Massachusetts, outlined the steps that he and his colleagues have taken since November to evaluate nearly 200 applicants for a Harvard faculty position. Interviews for six candidates begin this week. A tweet by Eddy on 9 January attracted fresh attention to the blog post, with commenters applauding his efforts to lift the veil on the selection process. Holly Bik, a genomics and bioinformatics researcher at New York University who is applying for jobs, tweeted: "Your blog post is a step in the right direction though; demystifying the black box process is always good."

➔ **NATURE.COM**
For more on popular papers:
go.nature.com/auyljg

ASTRONOMY

Brightest-ever supernova

A supernova has been spotted that is twice as luminous as the previous record holder — at its peak it blazed brighter than 570 billion Suns.

Subo Dong at Peking University in Beijing and his colleagues spotted the exploding star, called ASASSN-15lh, in June last year using an automated telescope survey of the sky. The object belongs to a class of 'superluminous' supernovae that are low in hydrogen. Whereas other hydrogen-poor supernovae reside in dim dwarf galaxies, this one seems to originate near the centre of a massive galaxy that is brighter than the Milky Way.

ASASSN-15lh challenges previous theories of explosion mechanisms and how supernovae are powered, the authors say.

Science 351, 257–260 (2016)

CONSERVATION

Protected areas miss key corals

Only a tiny percentage of the world's corals and tropical fish is safeguarded by current



marine protected areas.

David Mouillot at the University of Montpellier in France and his colleagues compared the geographical range of 805 species of reef-forming hard corals (pictured) and 452 tropical reef fish with a database of more than 3,600 marine protected areas. They then analysed the trees of life of these organisms and found that just 1.7% of the tree's 'branches' for the corals and 17.6% for the fishes had 10% or more of their ranges within the marine protected areas.

Improved conservation strategies are needed to better secure the biodiversity of these animals, the authors say. **Nature Commun.** 7, 10359 (2016)

➔ **NATURE.COM**
For the latest research published by Nature visit:
www.nature.com/latestresearch

SEVEN DAYS

The news in brief

EVENTS

Clinical-trial death

Five people were hospitalized and one has died after a French clinical trial went disastrously wrong. The phase I trial — designed to test the safety of new treatments in healthy people — involved a drug made by the company Bial near Porto, Portugal, and was conducted by a French contract-research organization, Biotrial, in Rennes. Neither the French authorities nor Biotrial has identified the drug, but Bial says that it was an inhibitor of the enzyme FAAH (fatty acid amide hydrolase). See page 263 for more.

Ebola returns

Health officials confirmed on 15 January that a 22-year-old woman in Sierra Leone had died of Ebola. The announcement comes two months after the World Health Organization declared that the disease had stopped spreading in Sierra Leone, and less than a day after it was announced that a similar resurgence of the virus in Liberia had ended. The source of the latest case is being investigated; other flare-ups have been traced to survivors who still harboured the virus in semen and other bodily fluids.

Turkish arrests

Twenty-seven Turkish academics were arrested, and later released, on 15 January

NUMBER CRUNCH

\$4.5 bn

Annual amount needed to fight pandemics, says a group convened by the US National Academy of Medicine.

Source: Commission on a Global Health Risk Framework for the Future



SPACEX

Launch was a success, landing less so

A joint US–European satellite that will track the surface height of the oceans was launched into orbit on 17 January. The oceanography mission Jason-3 will provide data for weather, climate and ocean researchers. But the

Falcon 9 rocket that delivered the satellite into orbit had a less happy outcome. It was supposed to touch down on an ocean barge, but instead toppled over and was destroyed on landing.

for signing a petition calling on their government to end violence in Turkey's southeast, where government forces have been fighting Kurdish separatists. The researchers face prosecution for alleged defamation of the state and spreading of terrorist propaganda. Around 2,000 scientists from about 90 Turkish universities have signed the petition. Several of the universities have launched investigations into signatories in their faculty. See go.nature.com/gpcarv for more.

Zika virus

The US Centers for Disease Control and Prevention has issued a travel alert over the ongoing transmission of Zika virus in 14 countries in the Caribbean and in

Central and South America. The alert — issued on 15 January — recommends that pregnant women consider postponing travel to any areas where Zika infection is occurring. It comes after evidence emerged that, since October, around 3,530 babies have been born in Brazil with unusually small heads and brains — a disorder called microcephaly. That is ten times more than the country usually sees in a year, and cases are concentrated in regions with Zika outbreaks.

BUSINESS

NASA spaceplane

Starting in 2019, NASA will gain a third commercial partner for flying cargo to and from the International Space

Station. The space agency announced on 14 January that it will use a miniature spaceplane owned by the Sierra Nevada Corporation in Sparks, Nevada, for a minimum of six missions by 2024. The craft can transport both pressurized and unpressurized loads, and lands on a runway back on Earth, improving NASA's options for delivering and returning fragile scientific equipment. NASA also uses spacecraft from Orbital ATK in Dulles, Virginia, which burn up on re-entry, and from SpaceX in Hawthorne, California, which splash down into the ocean.

Vaccine deal

Gavi, the international vaccine alliance, announced on 20 January that it has

paid US\$5 million to Merck, manufacturer of the first Ebola vaccine shown to protect against the virus in a human clinical trial. The deal marks the first time that the public-health organization has moved to purchase a vaccine before it has been licensed. In return for the payment, Merck promises that it will seek to have the vaccine approved by a regulatory agency by 2017. See go.nature.com/ujcirz for more.

PEOPLE

Sentence upheld

A former biomedical researcher at Iowa State University has failed to reduce his prison sentence for making false statements on successful funding applications to the US National Institutes of Health. On 11 January, a federal appeals court ruled against reducing the sentence of Dong Pyou Han, who pleaded guilty in 2015. At the time, Han was sentenced to 57 months in prison, but he appealed against the sentence two weeks later and argued that its length was unreasonable given his remorse and lawful past. The appeals court disagreed and upheld the sentence.

Taiwanese election

An epidemiologist famed for his work on the 2003 outbreak of severe acute respiratory syndrome (SARS) and for his



studies on the health burden of arsenic and hepatitis is to become Taiwan's vice-president. Chen Chien-Jen assumes the role after Tsai Ing-Wen, his running mate, won the 16 January election. Chen is widely respected for his work as health minister in Taiwan during the SARS epidemic, and was vice-president of the nation's premier research organization, the Academia Sinica. See go.nature.com/wrrwn for more.

Perpetrator named

Nature and other sources confirmed last week that the faculty member suspended by the California Institute of Technology for harassing two female graduate students is theoretical astrophysicist Christian Ott. Caltech, in Pasadena, announced on 4 January that it had suspended an unnamed faculty

member without pay for an academic year. Meanwhile, in a 12 January statement to the US House of Representatives, Congresswoman Jackie Speier (Democrat, California) highlighted sexual harassment in astronomy. She has called for universities to share information with each other on the outcome of harassment investigations. See go.nature.com/mnwdim for more.

POLICY

US coal pause

The United States has halted any new coal mining on federal land as it begins a major review of the coal industry. Interior secretary Sally Jewell announced the review on 15 January and said it would ensure that the federal coal programme took into account the impact of the industry on climate change. Coal power is one of the most greenhouse-gas-intensive forms of electricity generation. The halt on new leases applies while the comprehensive review — which Jewell says is the first in 30 years — is under way.

Iran sanctions

The United States and the European Union lifted a broad array of sanctions against Iran on 16 January after it was confirmed that the country had taken steps to impair its

ability to produce plutonium and enriched uranium, which are used in atomic bombs. The lifting of restrictions came within hours of confirmation by the International Atomic Energy Agency that Iran was in compliance with the deal it struck with six world powers in July over its nuclear programme. Sanctions have crippled Iran's economy and contributed to the international isolation of its scientists.

EU reprimand

The ombudsman for the European Union (EU), Emily O'Reilly, has ruled that delays by the European Commission in dealing with 20 authorization applications for genetically modified (GM) foods and feedstuffs were a "systemic problem" and "constituted maladministration". The ombudsman, who investigates complaints about EU bodies, said on 15 January that the commission had consistently failed to make decisions on the applications within the three-month deadline required by law. A review of EU decision-making on GM foods and feedstuffs is under way, and so O'Reilly did not make a recommendation related to the case. But she said that the commission should comply with current requirements until the review is complete.

Reproducibility call

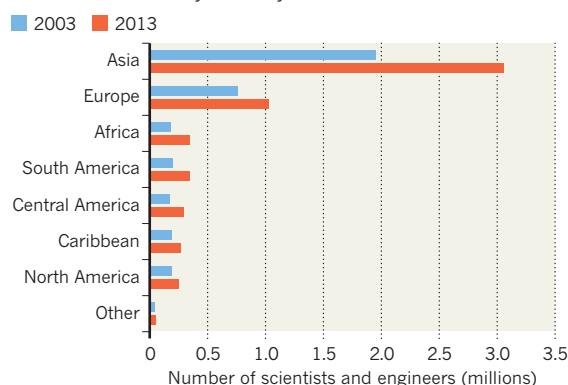
A coalition representing more than 100,000 experimental biologists has released recommendations for enhancing the reproducibility of research. The Federation of American Societies for Experimental Biology, which includes 30 scientific groups, wants to see better characterization of antibodies, easier reporting of negative results, better training for graduate students and established scientists, and fuller descriptions of animal models. See page 256 for more.

TREND WATCH

As the number of scientists and engineers in the United States jumped between 2003 and 2013, so too did the proportion of immigrants. A report from the US National Science Foundation (see go.nature.com/9hhf15) shows that immigrants made up just 15% of the 21.6-million-strong science workforce in 2003, rising to 18% of 29 million people in 2013. Asia remains the most likely region of birth for the immigrants; the number of Indian-born scientists nearly doubled, from 515,000 in 2003 to 960,000 in 2013.

IMMIGRATION INCREASES IN US SCIENCE

Asia still sends more scientists to the United States than does any other region. But all parts of the world are sending more scientists to the country than they were in 2003.



NEWS IN FOCUS

GENE-EDITING Patent dispute over CRISPR tool intensifies **p.265**

SPACE Best evidence yet for ninth solar system planet **p.266**

CLIMATE Research blitz aims to unlock secrets of El Niño **p.267**

GENETICS What happens when domestic fowl go wild in Hawaii **p.270**



STEPHANE MAHE/REUTERS



The hospital in Rennes, France, to which six people were taken after suffering adverse effects in a phase I clinical trial.

PHARMACOLOGY

Scientists in the dark after fatal French clinical trial

Knowledge about the drug's structure would help researchers understand what happened.

BY DECLAN BUTLER & EWEN CALLAWAY

One person died, and five others were hospitalized, after a clinical trial of an experimental drug in France went tragically wrong. But days after the first public acknowledgement of the incidents on 15 January, a lack of official information has left outside experts and the public largely in the dark as to what happened.

"The French authorities have not been very rapid nor transparent in their response," says Catherine Hill, a specialist in clinical-trial design and a former member of the scientific advisory board of France's National Agency for

Medicines and Health Products Safety (ANSM). She adds that French investigations into other medical accidents have often been opaque.

The trial was a 'first-in-human' phase I trial to test the drug's safety in healthy people (see 'Basic facts'). The Portuguese company Bial produced the drug, which was aimed at treating anxiety and motor disorders associated with Parkinson's disease, and chronic pain in people with cancer and other conditions. Biotrial, a French contract-research organization, conducted the trial at its facilities in Rennes.

But many key questions remain unanswered, says Marc Rodwin, a biomedical-law specialist at Suffolk University Law School in Boston,

Massachusetts. This includes how the participants' injuries came about — magnetic-resonance-imaging scans showed dying and bleeding tissue deep in the brain — and whether the trials were conducted properly.

BRAIN ENZYME

In particular, neither the French authorities nor Biotrial has disclosed the identity of the molecule administered in the trials. Bial did say that the drug was an FAAH (fatty acid amide hydrolase) inhibitor; FAAH is an enzyme produced in the brain and elsewhere in the body that breaks down neurotransmitters known as endocannabinoids. By ►

FATAL DRUG TRIAL

Basic facts

- The trial recruited 128 healthy volunteers aged 18–55, who were paid €1,900 (US\$2,060) each.
- Ninety people received different doses of the drug, and the remainder a placebo.
- The trial had tested escalating single doses of the drug without observing any serious adverse side effects.
- The six participants who fell ill were the first to receive repeat higher doses over the course of several days.
- The first participant to fall ill experienced adverse symptoms on 10 January and died

on 17 January.

- Biotrial halted the trial on 11 January; the other five affected people were hospitalized in the days that followed.
- One of these patients has since been discharged, and the condition of the other four is judged to be serious but stable.
- Authorities are contacting the 84 other people who received the drug at lower doses to arrange medical check-ups; none of the 18 given neurological check-ups over the weekend showed any of the symptoms seen in the hospitalized people.

► blocking these enzymes, FAAH inhibitors cause endocannabinoids — which activate the same neural receptors as the active chemical in cannabis, and might have painkilling properties — to accumulate in the body.

Some scientists scrambled over the weekend to try to establish the identity of the drug. Among them were Steve Alexander, a molecular pharmacologist at the University of Nottingham Medical School, UK, who has worked on FAAH for 15 years, and his colleague Christopher Southan, a curator for the Guide to Pharmacology database at the University of Edinburgh, UK. Together, the pair examined an online list of drugs in Bial's research pipeline.

The search revealed just two molecules in phase I trials, one of which fitted the therapeutic profile mentioned by Bial, although it was referred to only by a codename, BIA 10-2474. A French newspaper also published a recruitment form given to a volunteer in the trial that mentioned a drug with the same codename. "As best as we can make out, this compound has not been described in the [scientific] literature," says Alexander. "So we're working in the dark."

It is common in the pharmaceutical industry not to reveal the structure of a molecule this early in development — although the practice has been criticized by researchers. "They declare codenames of candidates in development and hide the structure," says Southan. "I think it's time they stopped." That lack of information left researchers trying to guess the structure from published Bial patents over the

weekend, Southan adds. He also says that there seems to be no entry for the trial in clinical-trial registries.

Numerous companies have developed FAAH inhibitors. There is none on the market, because most clinical trials have shown them to be ineffective — but the ones that were previously tested in people proved safe.

OFF-TARGET ACTION

Many researchers believe that BIA 10-2474 is acting 'off target' — in other words, inhibiting a protein other than an FAAH. To investigate, researchers could radioactively label the compound and test it on brain tissue from cadavers to 'fish out' the proteins it binds to.

Knowing the drug's molecular structure would also enable scientists to run computer predictions of this and other mechanisms that might result in toxicity. "There's a whole gamut of sophisticated computation analysis to predict anything you like," says Southan.

Other researchers studying the FAAH pathway will probably look more closely at the potential for inhibitors to strike other proteins, Alexander says. "I think it's very likely that both private industry and academic institutions will be looking very hard as to what this off-target affect might be."

The lack of transparency is typical of French investigations, which tend to favour secrecy until firm conclusions are established, says a French health-law specialist who requested anonymity. He notes that the country's rules governing research on human subjects are strong and guarantee substantial protection of trial participants. He adds that safety incidents in clinical trials are almost unheard of in the country, with the price often being delays in the approval of trial applications.

In recent years, there have been two major changes to French laws affecting the approval of drugs in clinical trials. France strengthened its medical-safety laws following the 2009 withdrawal of a diabetes drug that was suspected of causing hundreds of deaths: a 2011 law, in particular, tightened rules on conflicts of interest for people involved in the country's drug-approval process, as well as giving authorities more power to demand safety tests of medications after they are approved. Then, in 2012, the government passed a separate law intended to streamline the rules for research involving humans, to speed up therapeutic progress and to make France a more attractive place for companies to carry out clinical trials.

One possible safety issue in the trial of BIA 10-2474, notes trial-design specialist Hill, is that all six participants seem to have been administered the doses simultaneously, rather than one receiving a test dose and being checked for adverse effects before others were given it.

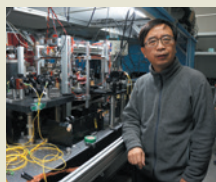
Simultaneous rather than sequential administration was identified as problematic in a disastrous UK clinical trial in 2006 that caused multiple organ failure in six participants. "From the 2006 catastrophe in London, I had concluded that treating several individuals with the same dose on the same day in a phase I trial was a big mistake," says Hill.

Jean-Marc Gandon, the president and chief executive of Biotrial, says that he cannot immediately respond to queries from *Nature*, that he is focused on trying to save the patients and that the company will respond later.

Bial spokeswoman Susana Vasconcelos says that the trial had been conducted "in accordance with all the good international practices guidelines, with the completion of tests and preclinical trials" and that the company "is committed to determine thoroughly and exhaustively the causes which are at the origin of this situation". ■



Q&A



China's quantum space pioneer, Pan Jian-Wei
go.nature.com/fftxy3

MORE NEWS

- Brightest-ever supernova still baffles astronomers go.nature.com/2vygzd
- Seven lessons from the Ebola epidemic go.nature.com/oenddk
- Earliest volcanic-eruption painting in iconic Chauvet cave go.nature.com/w6wgil

NATURE PODCAST



Degradable brain sensors, a 10,000-year-old murder mystery and feral chickens
nature.com/nature/podcast

TENG YUN CHEN

BIOTECHNOLOGY

Bitter fight over CRISPR patent heats up

Unusual battle among academic institutions holds key to gene-editing tool's future use.

BY HEIDI LEDFORD

A versatile technique for editing genomes has been called the biggest biotechnology advance since the polymerase chain reaction (PCR), and the US Patent and Trademark Office (USPTO) is set to determine who will reap the rewards.

On 11 January, the USPTO granted a request to review a key patent awarded for the technique, known as CRISPR–Cas9. The outcome of the ensuing proceedings, called a patent interference, could be worth millions to the research institutions that are at war over the relevant patents. It might also influence who is allowed to use the technology — and under what terms.

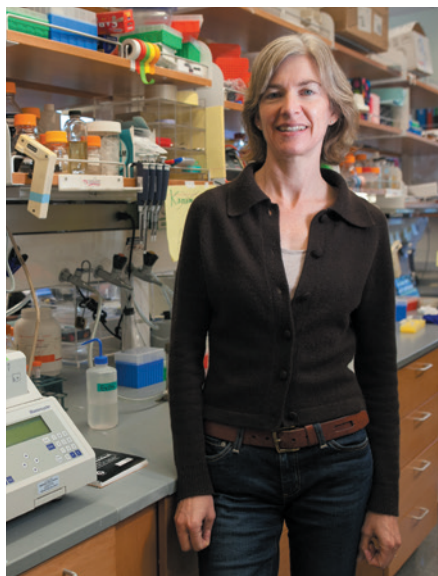
“This is an absolutely humungous biotech patent dispute,” says legal scholar Jacob Sherkow of New York Law School. “We’re all waiting with bated breath.”

CRISPR–Cas9 is a bacterial defence system that uses the enzyme Cas9 to snip DNA at sites determined by the sequence of a ‘guide’ strand of RNA. The technology allows scientists to disable, replace or tweak genes by rewriting snippets of DNA sequences. Use of CRISPR–Cas9 in research has exploded, thanks to its relative simplicity and versatility compared to other gene-editing methods. Several companies have sprung up to harness the technique for generating improved crops, research reagents and therapies for human genetic diseases.

The roots of the CRISPR–Cas9 patent fight date back to 2012, when researchers reported that they had reprogrammed the system to cut strands of isolated DNA at sites of their choosing¹. The team, led by biologists Jennifer Doudna at the University of California, Berkeley, and Emmanuelle Charpentier, now at the Max Planck Institute for Infection Biology in Berlin and Umeå University in Sweden, filed a patent application on 15 March 2013.

By then, publications had emerged from other groups showing that the method works in human cells^{2–4} — bolstering dreams of CRISPR-based gene therapies.

One of those groups, led by synthetic biologist Feng Zhang of the Broad Institute and the Massachusetts Institute of Technology, both in Cambridge, filed a patent application for the CRISPR–Cas9 technique in October 2013.



Jennifer Doudna of the University of California, Berkeley, helped to develop the CRISPR system.

The institutions filed the patent under a special expedited review programme, and it was granted in April 2014. Zhang has since been awarded additional patents on the technology. The original Doudna–Charpentier patent remains under review.

A PITCHED BATTLE

In April 2015, the Berkeley team asked the USPTO to determine which team was the first to invent the technique. Now the patent office has agreed to begin that interference proceeding, which will be much like a court case, with both sides presenting evidence culled from publications and laboratory notebooks. “Once the [USPTO] declares an interference, that’s really when the fur is going to fly,” Sherkow predicted in a June interview.

The patent interference is also a testament to the high stakes involved: companies aiming to use CRISPR–Cas9 for gene therapy have raised hundreds of millions in venture capital and other funds in less than three years. One company, Editas Medicine in Cambridge, Massachusetts, has already filed to go public.

Arti Rai, a legal scholar at Duke University in Durham, North Carolina, says that it is unusual for academic research institutions to battle so intensely over a patent. Instead, such

institutions usually come to an agreement to share rights to the invention. “This seems more bitter than disputes I’ve heard of in the past,” she adds.

The two patents in question make broad claims to ‘foundational’ intellectual property thought to be necessary for most lucrative CRISPR–Cas9 applications. But many patents have been filed on CRISPR–Cas9 technologies, and there is still the chance that the winner of the interference will face additional challenges in court. Zhang’s group has also reported another enzyme, called Cpf1, that functions much like Cas9. Researchers expect other alternatives to emerge with time.

LICENSING LOOMS

For now, it is unclear how the dispute will affect researchers who use CRISPR–Cas9, if it does so at all. “Patent holders might send out a few cease-and-desist letters, but they probably won’t sue academic researchers,” says Rodney Sparks, a biotechnology-patent counsel at the University of Virginia in Charlottesville. Doing so would take time and money with little reward: the spoils in a patent lawsuit are typically damages or a share of royalties from a marketed product. That leaves little to gain from suing academics who are not selling anything. But those who intend to use their research as the basis for a start-up company will need to be wary, Sparks says.

Some patent holders do ask that even scientists doing basic research take out a licence on a patented technology, typically for a fairly small fee. Such was the case for PCR, says Warren Woessner, a lawyer at Schwegman, Lundberg and Woessner in Minneapolis, Minnesota. The Broad Institute has noted on its website that it will continue to make CRISPR–Cas9 reagents available to the community, and has given no indication that it will pursue hefty licensing fees from academics. But Sherkow warns against assuming that the spirit of academic camaraderie will always prevail: licensing revenue has become increasingly important, particularly for major research institutions, he says. “We’re just living in a brave new world these days.” ■

1. Jinek, M. *et al. Science* **337**, 816–821 (2012).
2. Cong, L. *et al. Science* **339**, 819–823 (2013).
3. Mali, P. *et al. Science* **339**, 823–826 (2013).
4. Jinek, M. *et al. eLife* **2**, e00471 (2013).

Unseen planet may lurk near Solar System's edge

Gravitational signature hints at massive object that orbits the Sun every 20,000 years.

BY ALEXANDRA WITZE

A century after observatory founder Percival Lowell speculated that a 'Planet X' lurks at the fringes of the Solar System, astronomers say that they have the best evidence yet for such a world. They call it Planet Nine.

Orbital calculations suggest that Planet Nine, if it exists, is about ten times the mass of Earth and swings an elliptical path around the Sun once every 10,000–20,000 years. It would never get closer than about 200 times the Earth–Sun distance, or 200 astronomical units (AU). That range would put it far beyond Pluto, in the realm of icy bodies known as the Kuiper belt.

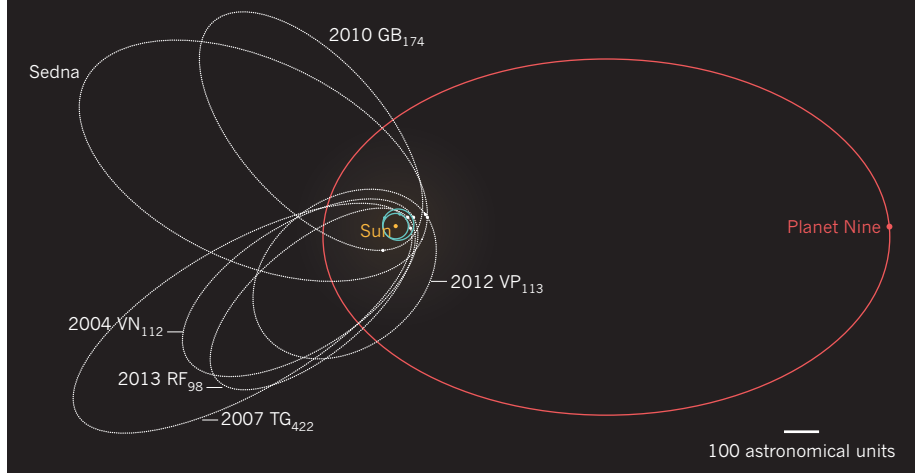
No one has seen Planet Nine, but researchers have inferred its existence from the way several other Kuiper belt objects (KBOs) move. And given the history of speculation about distant planets (see 'Solving for X'), Planet Nine may end up in the dustbin of good ideas gone wrong.

"If I read this paper out of the blue, my first reaction would be that it was crazy," says Mike Brown, an astronomer at the California Institute of Technology in Pasadena who was part of the research team. "But if you look at the evidence and statistics, it's very hard to come away with any other conclusion."

Brown and his colleague Konstantin Batygin propose Planet Nine in a paper published on 20 January in the *Astronomical Journal*

FAR AFIELD

The existence of an unseen 'Planet Nine' could explain the strange orbits of several objects (whose orbits are shown in white) in the Kuiper belt beyond Neptune.



(K. Batygin and M. E. Brown *Astronom. J.* 151, 22; 2016).

Alessandro Morbidelli, an orbital-dynamics specialist at the University of the Côte d'Azur in Nice, France, who has reviewed the paper in detail, says he is "quite convinced" that the planet exists. Others are not so sure.

"I have seen many, many such claims in my career," says Hal Levison, a planetary scientist at the Southwest Research Institute in Boulder,

Colorado. "And all of them have been wrong."

Claims of Planet Nine's existence recall a period in the nineteenth century when astronomers predicted and then discovered Neptune by studying tiny perturbations in the orbit of Uranus. The gravity of some unseen body must be tugging on Uranus, they said — and they were right. "In some sense we're hoping to relive history a little bit," says Batygin.

The story of Planet Nine began in 2014,

SOLVING FOR X

Astronomers have long speculated about the existence of additional large planets in the outer Solar System, but none has yet been confirmed.

1846 Johann Gottfried Galle discovers Neptune, guided by predictions from perturbations of Uranus's orbit.

1905 Percival Lowell (pictured) starts hunting for a 'Planet X', which he predicted would lie beyond Neptune, just as Neptune lies beyond Uranus. His calculations led astronomers at Lowell's namesake observatory to find Pluto in 1930, but the object is not massive enough to be Planet X.



1984 On the basis of periodic extinctions in the fossil record, scientists propose that a dwarf star, later named Nemesis, passes through the Solar System every 26 million years, flinging comets on a path to impact Earth.

1999 Perturbations in comet orbits lead astronomers to propose that a brown dwarf (bigger than a planet but smaller than a star) exists in the outer Solar System. It is named Tyche, the good sister of Nemesis.

2014 A search with the Wide-Field Infrared Survey Explorer satellite rules out the existence of both Nemesis and Tyche. But the discovery of an object in the distant Kuiper belt prompts Chadwick Trujillo and Scott Sheppard to propose a large planet in the Kuiper belt.

2016 Orbital calculations by Konstantin Batygin and Mike Brown strengthen the concept of this unseen planet, which they name 'Planet Nine'.

K. BATYGIN AND M. E. BROWN *ASTRONOM. J.* 151, 22 (2016)

EVERETT COLLECTION HISTORICAL/ALAMY

when a pair of astronomers reported finding a KBO called 2012 VP₁₁₃. Its stretched-out orbit never came closer than 80 AU to the Sun (C. A. Trujillo and S. S. Sheppard *Nature* **507**, 471–474; 2014). (Pluto, at its most distant, is 48 AU from the Sun.) VP₁₁₃ joined the dwarf planet Sedna as only the second known object with a very distant orbit. In their report, Chadwick Trujillo at the Gemini Observatory in Hilo, Hawaii, and Scott Sheppard of the Carnegie Institution for Science in Washington DC said that the orbits of these objects suggested that yet another object, a planet bigger than Earth, could exist at around 250 AU (see ‘Far afield’).

Batygin and Brown picked up the challenge. “Our main goal at that point was to show that this idea is crazy,” says Brown.

But Trujillo and Sheppard had noted that Sedna, VP₁₁₃, and several other KBOs all shared a peculiar property: their closest approach to the Sun lay in the plane of the Solar System, and they all moved from south to north when crossing that plane.

Batygin and Brown analysed the orbits further and discovered that their long axes were physically aligned, too, as if something had nudged them to occupy the same region of space around the Sun. The team concluded that a massive object must be shepherding the objects. “We have a gravitational signature of a giant planet in the outer Solar System,” Batygin says.

Planet Nine — informally known as Phat-tie — is probably smaller than Neptune and icy with a gassy outer layer. The gravitational effect of Uranus and Neptune would have flung it outward in the first 3 million years of the Solar System’s existence, Batygin says.

Actually spotting Planet Nine through a telescope could be difficult because it would spend most of its time very far from the Sun, making it faint and hard to see, notes Meg Schwamb, an astronomer at the Academia Sinica in Taipei. Brown and Batygin have been looking for it using the Subaru telescope in Hawaii, so far without success. The Large Synoptic Survey Telescope in Chile will have a good chance of catching it when it starts operating early next decade, Brown says.

But he and Batygin say that there are other ways to test the existence of Planet Nine. Its gravitational influence would also produce a population of KBOs with orbits at steeply inclined angles. A few of these have already been spotted, but discovering more would strengthen the statistics of the discovery and help to clarify whether Planet Nine really exists or not, says David Nesvorný, a planetary scientist at the Southwest Research Institute. So it’s back to the telescopes. “It really points to the fact that more extreme KBOs need to be found,” says Trujillo. “The location is not known well enough to just point a telescope at it and say, ‘there it is.’” ■



El Niño has sent heavy rains to California this year.

CLIMATE SCIENCE

Meteorologists probe monster El Niño

Unprecedented Pacific Ocean campaign aims to improve forecasts for strong storms.

BY RICHARD MONASTERSKY

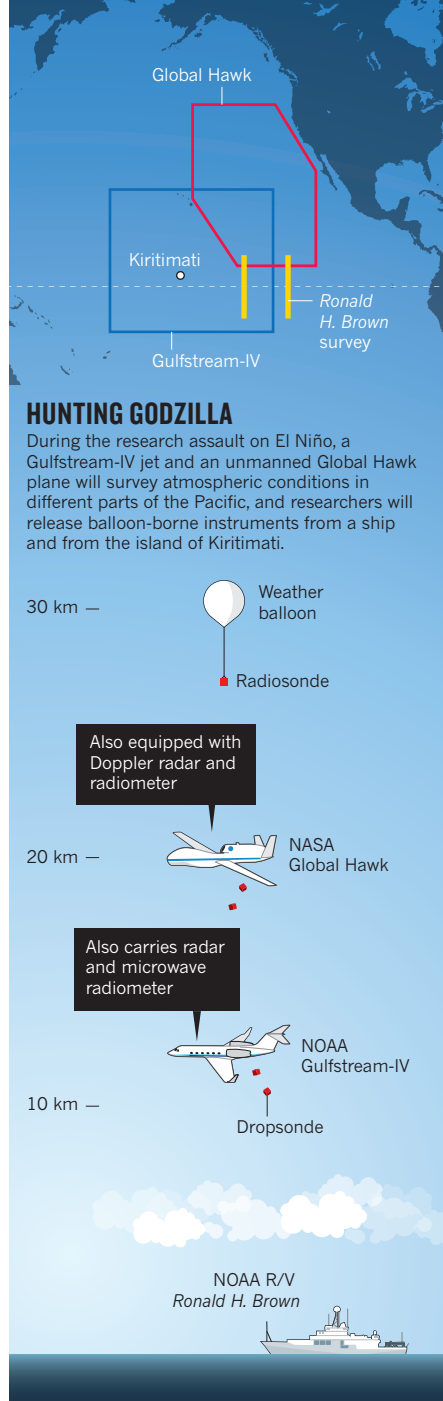
Climate scientists this week began a research blitz to study El Niño, the climate trouble-maker that disrupts weather around much of the globe. For the next two months, US researchers will use specially outfitted planes, a research ship and hundreds of weather balloons to monitor the region in the tropical Pacific Ocean where El Niño forms. Ultimately, the scientists say, their measurements could help to improve weather forecasts and unlock secrets about how powerful El Niño events evolve.

“We’re seeing an extreme climate state — one that we know tends to produce extreme climate conditions worldwide,” says Randall Dole, lead scientist on the project and a meteorologist with the US National Oceanic and Atmospheric Administration (NOAA) in Boulder, Colorado. “We’re going

right into the heart of that.”

The El Niño warming that has bloomed in the equatorial Pacific is one of the strongest on record, with ocean temperatures reaching as much as 3 °C above normal across the central and eastern parts of the ocean basin. That heat triggers convection in the atmosphere that re-routes major air currents, often sending strong storms towards California while drying out parts of southeast Asia, Australia and eastern South America. But researchers have few data on the atmospheric changes in the core of the El Niño region because the remote equatorial Pacific is essentially a meteorological black hole.

To begin the campaign, NOAA sent its Gulfstream-IV research jet to Hawaii, its base for about 20 flights south towards the Equator. Using onboard remote-sensing equipment and dropsondes — packages of instruments released from the plane — the team will measure winds, temperature, ►



HUNTING GODZILLA

During the research assault on El Niño, a Gulfstream-IV jet and an unmanned Global Hawk plane will survey atmospheric conditions in different parts of the Pacific, and researchers will release balloon-borne instruments from a ship and from the island of Kiritimati.

► air pressure and moisture from a height of 12–14 kilometres down to the ocean surface (see ‘Hunting Godzilla’).

In February, NASA’s unmanned Global Hawk aircraft will join the effort, prowling the eastern part of the Pacific in 4 flights lasting up to 24 hours each. At the same time, NOAA will launch instrument packages on weather balloons from Kiritimati, or Christmas Island, an atoll near the Equator in the heart of the region in which El Niño forms. And researchers will also release balloon-borne instruments from the NOAA research ship *Ronald H. Brown* as it conducts a previously planned cruise in the central Pacific.

SEIZING THE MOMENT

The idea for the roughly US\$3-million campaign developed as the warming gathered strength last year; Dole and his colleagues realized that they had a rare opportunity to collect the first detailed atmospheric measurements of a monster El Niño. NOAA scrambled to pull the campaign together in a few months — rather than the usual two to three years that it usually takes to mount a major meteorological field project.

The agency had some resources to spare: thanks to the way El Niño alters conditions over the Atlantic, there were relatively few tropical storms there last year. That meant that NOAA did not use all of the flying time budgeted for the Gulfstream-IV hurricane hunter, which flies over storms to collect data useful for forecasters. The quiet hurricane season also meant that the Global Hawk did not make as many research flights in the Atlantic as planned last year.

“We’ve done this largely by reallocating,” says Dole. “We’re working within the existing budget and shifting everything around.”

Alexey Fedorov, a climate modeller at Yale University in New Haven, Connecticut, says that because extreme El Niño events are so rare, “it is important to use any opportunity to

gather as much data as possible”. Fedorov, who is not part of the campaign, says that researchers lack a full understanding of the way that strong El Niños evolve and alter global weather patterns.

The information gathered over the next few months could yield long-term dividends for El Niño researchers, says Dole. “If we do this well, it will impact our community for the next 10 or 20 years.”

But the project’s immediate goal is to help forecasters to understand how the unruly atmosphere will affect weather now. By gathering direct measurements from this data-poor zone, leaders of the NOAA campaign hope to improve weather forecasts and allow researchers to test weather models to better understand the source of errors in those models.

Data from the Global Hawk will also aid meteorologists tracking El Niño-spawned storms as they barrel down on the western

“If we do this well, it will impact our community for the next 10 or 20 years.”

United States, says Dole. Over the past few weeks, coastal California has been pummelled by such storms, and more are expected. As part of the El Niño

campaign, NOAA has installed a scanning X-band radar south of San Francisco Bay that will measure precipitation in approaching storms.

The agency will upload data from the field campaign to the World Meteorological Organization’s Global Telecommunications System, so that forecasters around the globe can access the observations. Peter Bauer, an atmospheric modeller at the European Centre for Medium-Range Weather Forecasts in Reading, UK, says that he plans to feed the data into model experiments with the aim of improving forecasts for Europe. The campaign, he says, “has potentially a very big impact”. ■

RESEARCH INTEGRITY

GM-crop papers spark probe

Work describing harm from genetically modified crops was cited in Italian Senate hearing.

BY ALISON ABBOTT

Papers that describe harmful effects to animals fed on genetically modified (GM) crops are under scrutiny for alleged data manipulation. The leaked findings of an ongoing investigation at the University of Naples in Italy suggest that images in the papers may have been intentionally altered.

The leader of the lab that carried out the work says that there is no substance to this claim.

The papers’ findings run counter to those of numerous safety tests carried out by food and drug agencies around the world, which indicate that there are no dangers associated with eating GM food. But the work has been widely cited on anti-GM websites — and results of the experiments that the papers describe were

referenced in an Italian Senate hearing last July on whether the country should allow cultivation of safety-approved GM crops.

“The case is very important also because these papers have been used politically in the debate on GM crops,” says Italian senator Elena Cattaneo, a neuroscientist at the University of Milan whose concerns about the work triggered the investigation.

Following the Senate hearing, Cattaneo took a closer look at three papers¹⁻³, which all emerged from a research lab at the University of Naples, headed by veterinary scientist Federico Infascelli. They describe experiments on goat kids born to mothers fed on GM soya-bean meal, and conclude that fragments of the foreign gene in the soya bean can be transported across the gut and secreted in the milk, influencing the biology of the suckling kids.

Cattaneo noted what looked like problems in all three papers: sections of images of electrophoresis gels appeared to have been obliterated, and some of the images in different papers appeared to be identical but with captions describing different experiments.

She commissioned Enrico Bucci, head of biomedical-services firm BioDigitalValley in Aosta, Italy, to carry out a forensic analysis of the papers. The analysis suggested that the papers indeed contained manipulated and reused images. Cattaneo contacted the journals concerned in September last year, and in November she sent the analysis to the University of Naples. The university rector, Gaetano Manfredi, an engineer, launched the university investigation, which is nearly complete. He says that the university will probably announce any resulting actions by the end of February.

However, details of the confidential findings of the investigation committee — composed



GM soya bean has passed numerous safety tests.

of scientists in and outside of Naples — were leaked to the Italian press. Tommaso Russo, a molecular biologist at the University of Naples who is coordinating the investigation, told *Nature* that the committee has found that the papers contain intentional data manipulation.

But, according to the Italian newspaper *La Repubblica*, Infascelli said that there is no substance to these allegations, and that an expert that he consulted about the papers had ruled out the possibility of data manipulation.

Infascelli declined to discuss the case with *Nature* until the investigation is complete.

On 14 January, Bucci posted his analysis of the three papers, as well as of four more papers on GM feed that were co-authored by Infascelli, and of a PhD thesis from Infascelli's lab. The analysis claims evidence for image manipulation in all eight papers. Bucci has informed Manfredi and Infascelli of his findings.

Since Cattaneo sent her findings to the journals concerned, one of the three papers under investigation has been retracted. The paper², published in *Food and Nutrition Sciences*, was withdrawn with a citation of “self-plagiarism”. The journal noted that the results were still valid and that the issues were down to “honest error”.

Plant geneticist Wayne Parrott at the University of Georgia in Athens has independent concerns about image manipulation in the three papers, and in another from Infascelli's lab that claims⁴ a metabolic impact of GM food in rabbits, which is also mentioned in Bucci's analysis. He has notified the relevant journals.

The investigation at the University of Naples is the first to test formal rules on scientific misconduct that Manfredi introduced last July. ■

1. Tudisco, R. *et al. Animal* **4**, 1662–1671 (2010).
2. Mastellone, V. *et al. Food Nutr. Sci.* **4**, 50–54 (2013).
3. Tudisco, R. *et al. Small Rumin. Res.* **126**, 67–74 (2015).
4. Tudisco, R. *et al. Anim. Sci.* **82**, 193–199 (2006).



WHEN CHICKENS GO WILD

The feral chickens of Kauai provide a unique opportunity to study what happens when domesticated animals escape and evolve.

BY EWEN CALLAWAY

“Don’t look at them directly,” Rie Henriksen whispers, “otherwise they get suspicious.” The neuroscientist is referring to a dozen or so chickens loitering just a few metres away in the car park of a scenic observation point for Opaekaa Falls on the island of Kauai, Hawaii.

The chickens have every reason to distrust Henriksen and her colleague, evolutionary geneticist Dominic Wright, who have travelled to the island from Linköping University in Sweden armed with traps, drones, thermal cameras and a mobile molecular-biology lab to study the birds.

As the two try to act casual by their rented car, a jet-black hen with splashes of iridescent green feathers pecks its way along a trail of bird feed up

to a device called a goal trap. Wright tugs at a string looped around his big toe and a spring-loaded net snaps over the bird. After a moment of stunned silence, the hen erupts into squawking fury.

Opaekaa Falls, like much of Kauai, is teeming with feral chickens — free-ranging fowl related both to the domestic breeds that lay eggs or produce meat for supermarket shelves and to a more ancestral lineage imported to Hawaii hundreds of years ago.

These modern hybrids inhabit almost every corner of the island, from rugged chasms to KFC car parks. They have clucked their way into local lore and culture and are both beloved and reviled by Kauai’s human occupants. Biologists, however, see in the feral animals an improbable experiment

RICH REID/NATL GEOGRAPHIC CREATIVE

A wild rooster in Kauai, Hawaii. in evolution: what happens when chickens go wild?

The process of domestication has moulded animals and their genomes to thrive in human environments. Traits that ensure survival in the wild often give way to qualities that benefit humans, such as docility and fast growth. Feralization looks, on its surface, like domestication in reverse. But closer inspection suggests that the chickens of Kauai are evolving into something quite different from their wild predecessors, gaining some traits that reflect that past, but maintaining others that had been selected by humans. In this way, they are similar to other populations of animals, including dogs, pigs and sheep, that have broken free of captivity and flourished.

By looking at feral animals, some evolutionary biologists hope to determine how domestic animals and their genes change in response to natural pressures. The research could also help to inform tricky conservation questions about how such animals affect native species, and ultimately whether and how to control them.

The natural history of the Kauai fowl makes them an important test case. “People have a really complicated relationship with the chickens,” says Eben Gering, an evolutionary ecologist at Michigan State University’s Kellogg Biological Station in Hickory Corners, who is in Kauai with Henriksen and Wright. “Some people absolutely want them gone. Some consider them an integral part of the local culture.”

FOUL-WEATHER FOWL

The Polynesian mariners who first settled the Hawaiian Islands about 1,000 years ago brought what they needed to start civilization anew. Staple plants such as taro, sweet potato and coconut palm made the trans-Pacific voyage, as did domestic dogs, pigs and, naturally, their prized chickens.

The Polynesian poultry probably bore little resemblance to the birds that provide much of the world’s protein today. Archaeological and genetic evidence suggests that they were more like red junglefowl (*Gallus gallus*) — small, furtive birds that still roam the forests of south-east Asia and are ancestral to all domestic chickens.

By the time that Captain James Cook landed in Waimea in southern Kauai in 1778, the Polynesian chickens had already, in essence, become feral. They wandered freely between Native Hawaiian villages and the neighbouring forest. Later, European and US settlers imported predators such as mongooses, which devastated birds of all kinds. Polynesian chickens were all but wiped out everywhere but Kauai and neighbouring Niihau, where the predators were never introduced.

On Kauai, chickens flourished. Although the birds’ numbers have not been tracked precisely, many residents contend that the population surged after hurricanes in 1982 and 1992 blew modern chickens from people’s back yards into the forests, where they encountered the descendants of the Polynesian chickens.

Gering says it is possible that interbreeding between the two populations allowed the birds’ numbers to swell. And during his first trip to the island in 2013, he and Wright noticed¹ that many of the feral chickens they encountered had flecks of the white feathers common in modern domestic breeds, in addition to the darker plumage usually seen in wild populations. Many had yellow legs (red-junglefowl legs are grey). And some of the roosters’ crows sounded conspicuously like the drawn out cock-a-doodle-doo of their barnyard brethren, rather than the truncated calls of red junglefowl.

DNA from 23 chickens revealed just how far domestic-chicken genes had infiltrated¹. The birds’ nuclear genomes seemed to be a mixture of genes from red-junglefowl-like Polynesian chickens and domestic chickens, whereas their maternally inherited mitochondrial markers traced back to European and Pacific domestic poultry. Gering and Wright think that a single hybrid population of feral chickens now roams Kauai, bearing a mixture of traits from modern and ancient birds.

In unpublished work, the two have scoured the birds’ genomes for stretches of DNA with very little variation across the population. This

homogeneity suggests that a gene has surged through the population in the recent past, probably because it offers some benefit. If feralization were domestication played backwards, then these ‘selective sweeps’ might appear around the DNA sequences that distinguish domestic chickens from red junglefowl. Instead, the researchers have found that most of the swiftest-evolving genes in the Kauai chickens are distinct from those suspected involved in modern domestication.

In some cases, genes from the Polynesian chickens are helping the hybrid feral chickens to adapt to Kauai’s habitats. For example, modern domestic fowl have been bred not to sit on, or brood, their eggs (making the eggs easier to collect). But in the wild, this trait puts unhatched chicks at risk. Wright and Gering found that feral chickens possess red-junglefowl gene variants that are linked to brooding.

But some genes of domestication do seem to be handy outside the coop. A variant linked to increased growth rates and reproduction in domestic chickens, for example, persists in the Kauai population, even though the average adult feral chicken is half the weight of a month-old bird bred for meat.

CHICKEN AND EGG

“You won’t see a bird as healthy-looking as that,” Wright says of the hen that he and Henriksen had captured at Opaekaa Falls. “Her plumage is perfect.” In the basement of a rented house on Kauai, the researchers have set up a makeshift laboratory where they photograph the bird, draw its blood and then kill it and prepare it for dissection. Wright starts with the hen’s Brazil-nut-sized brain.

Their unpublished research has shown that the brains of domestic chickens are smaller than those of junglefowl, relative to their body size, and organized differently. The team hopes to identify the genes responsible for these changes and others, such as the diminished visual-processing systems of domestic birds. Life in the wild has also altered the reproductive systems of the feral chickens. Domestic breeds lay eggs almost daily, but breeding seasonally could allow feral chickens to reapportion the minerals devoted to eggs (which come from spongy

“BEFORE DECIDING HOW IMPORTANT IT IS TO CONSERVE THEM, MANAGE THEM OR CULL THEM, IT WOULD AT LEAST BE GOOD TO KNOW ABOUT THEIR IMPACT.”

tissue in the centre of their bones) to making their skeletons more robust. The researchers sample the hen’s femur and also find that its ovaries are empty of egg follicles, which could be a sign of seasonal breeding.

Feralization has garnered much less attention from scientists than domestication (which gets a nod in chapter one of Charles Darwin’s 1859 *On the Origin of Species*). But swapping of domestic and wild genes has been happening all over the world for thousands of years. A feral-sheep population that has lived on the island of St Kilda in the Scottish Outer Hebrides for as long as 4,000 years acquired beneficial alleles that determine coat colour from a modern domestic sheep breed some 150 years ago². A 2009 study in *Science*³ found that some wolves in Yellowstone National Park, Wyoming, carry a domestic-dog version of a gene linked to dark coats that shows hallmarks of positive selection, possibly helping wolves from the Arctic to adapt to forested environments. “People would have thought that genes to live in a farm and house aren’t going to be any good in the wild, but that’s not necessarily true,” says Jonathan Losos, an evolutionary ecologist at Harvard University in Cambridge, Massachusetts.

And like Kauai’s feral chickens, other feral animals such as dingoes in Australia and urban pigeons practically everywhere have not evolved back to the state of their wild ancestors — even if certain traits may trend in that direction.

➔ NATURE.COM

To hear more about the chickens of Kauai, visit: go.nature.com/gnh7dj

FREE BIRD

The feral chickens of Kauai, Hawaii, are descendants of birds brought to the island by Polynesian settlers hundreds of years ago, which later mated with more modern domestic breeds. The chickens offer lessons in the genetics of feralization and on which traits tend to revert to a wild-like state.

COMB

Three of the fastest-evolving genes in the Kauai chicken are linked to comb size, a sexually selected trait that is not generally under pressure in domesticated breeds.

REPRODUCTION

Domestic egg-laying chickens do not sit on (or brood) their eggs, a trait selected by humans to make egg-collection simpler. Feral chickens favour gene variants associated with brooding, and the hens seem to lay eggs less frequently than domesticated birds — possibly seasonally.

SIZE

A gene variant linked to fast growth in domestic breeds persists in the Kauai chickens, yet a full-grown feral chicken is about half the size of a month-old domestic bird bred for meat.

FEATHERS

The plumage of the Kauai chickens is the most obvious sign of their mixed ancestry. White feathers, sported by many modern domestic breeds, compete with colours common to the red junglefowl (*Gallus gallus*) from which all chickens descend.



Feral chicken

Domestic chicken

LEGS

Domestic chickens have yellow legs, whereas those of the red junglefowl are grey, owing to variations in a gene called *BCDO2*. Both yellow- and grey-legged chickens roam Kauai.

WILD TO FERAL

Feralization is not domestication in reverse, even though it can look a bit like it. But feral and domesticated animals are known to breed with wild populations.



NO LOOKING BACK

Domesticated animals have gone feral all over the world. Interbreeding with wild animals can pose interesting questions for geneticists and headaches for conservationists. Human influences on these animals and on the environments they inhabit are probably intertwined, and almost definitely irreversible.



CANINE CONUNDRUM

Genes of domestication can carry advantages. Domestic-dog genes are helping North American wolves to adapt to a world with less tundra. In Yellowstone National Park, Wyoming, some grey wolves carry a dog gene variant linked to darker coats that may be beneficial in forests.



EQUINE EQUIVALENCE

The lines between wild, domestic and feral are fuzzy. Przewalski's horse on the Mongolian steppe has been touted as the last population of truly wild horses. But genome sequencing suggests that the group's ancestors routinely interbred with domestic horses.



SWINE SENSIBILITY

For some traits, domestication seems to be unidirectional. Feral pigs in Sardinia have regained the brain-size proportions and density of neurons involved in olfaction that are found in their wild relatives. But they nevertheless lack a molecule important for wild boars' exquisite sense of smell.

Like chickens, other domesticated animals tend to have smaller brains than their wild cousins, relative to body size (see 'Free bird'). And brain regions involved in processing things such as sight, sound and smell are among the most diminished, perhaps because humans bred animals to be docile and less wary of their surroundings. Feral pigs in Sardinia seem to have regained large brains and high densities of neurons involved in olfaction, but not the abilities that come with them: their neurons do not express a protein that has been linked to the exquisite sense of smell in closely related wild boars⁴. Likewise, feral dogs, cats and pigs often lack the savvy of their wild brethren and still depend on human niches for their survival, notes Melinda Zeder, an archaeologist at the Smithsonian Institution's National Museum of Natural History in Washington DC. Packs of feral dogs, for instance, do not form the complex hierarchies that make wolves such fearsome predators. "There's no leadership the way you get in a wolf pack. It's just a bunch of shitty friends," says Greger Larson, an evolutionary geneticist at the University of Oxford, UK, who is part of a team examining the mixed ancestry of Kauai's feral pigs.

Wright and Henriksen take less than an hour to dissect the captured hen and preserve samples of its brain, bone, liver and other tissue for gene-expression studies. They will use the RNA molecules expressed in different tissues to come up with a list of genes that might influence traits that distinguish the feral chickens from domestic birds and red junglefowl. They are eager to add to their study sample, and they jump at an offer to visit a nearby farm to collect more chickens.

PECKING ORDER

"They are a scourge. They are vermin. They cost us thousands and thousands of dollars every year," says the farm's owner (who asked not to be named). The birds scratch at tree saplings on his orchard, exposing the roots and killing fruit trees before they can mature. He patrols his property in a beaten-up luxury sedan with a high-powered air rifle and a hired hand who gets US\$5 per kill. Every few months, he invites hunters with night-vision goggles to visit and pick off the birds as they roost.

Few Kauaians share his malice towards feral chickens. Many locals give them a 'no-big-deal' shrug when asked. And the island's many tourists tend to view the birds with curiosity followed by mild annoyance after a couple of 3 a.m. wake-up calls. Feral-chicken merchandise — postcards, kitchen chopping boards, T-shirts — are ubiquitous. A popular children's television show hosted by a character called Russell the Rooster has been on air for nearly two decades.

As descendants of the birds imported by the Polynesians, Kauai's feral chickens occupy a zoological purgatory somewhere between native plants and animals and the dreaded invasive species that plague island habitats such as Hawaii. "It's much more complicated than just a feralized chicken," says Gering. "Even though junglefowl were not here before the Polynesians colonized the island, they have been a part of this ecosystem for much longer than the domestic chickens."

The chickens enjoy semi-official protection as 'wild chickens' in nature preserves. But if the same birds wander into developed areas or private property, they are considered 'free-flying domestic chickens' with no sanctuary. "Locals are free to take them (if they come onto your property) and put them in the pot," says a State of Hawaii website. (Gering and Wright, with a freezer full of the animals, consider making 'feral coq au vin'.)

Kauai may have no shortage of feral chickens now, but if mongooses arrive on the island or political winds change, they could be at risk. Hawaii's most populous island, Oahu, has mounted a controversial culling campaign against its feral chickens (whose ancestry is uncertain). But Gering thinks that the Kauai chickens' long tenure and unique cultural position makes some form of conservation worth considering. "Before deciding how important it is to conserve them, manage them or cull them, it would be good to at least know about their impact," he says. Researchers want to know about everything from the animals and plants that the birds eat to how they alter landscapes — information that Gering hopes to gather on future trips to the island.

Kauai's chickens are hardly the only creatures to occupy a nebulous space between native and alien. When Przewalski's horses, which live on the Mongolian steppe, were first described in the late nineteenth century, they were believed to be the planet's last wild, undomesticated horses. But a 2015 genome study⁵ found that the 2,100 or so remaining horses carry substantial amounts of domestic-horse DNA. They also show significant signs of inbreeding, owing to a captive-breeding programme begun in the 1940s.

"WHAT FERAL ANIMALS MAKE US DO IS CONSIDER THIS ALL-TOO-EASY DICHOTOMY BETWEEN WILD AND DOMESTIC."

Some conservationists view domestic genes as pollutants that are tarnishing the genomes of wild animals such as wolves, coyotes and even the red junglefowl native to southeast Asia. Some even contend that there are no 'pure' red junglefowl left.

"What feral animals make us do is reconsider this all-too-obvious, all-too-easy and all-too-wrong dichotomy between wild and domestic," says Larson.

And the ability of supposedly wild animals to thrive in a world increasingly altered by human activity may be due in part to the domestic genes that they now carry. What better way to adapt to human-moulded environments than to borrow traits from human-moulded creatures?

HOME TO ROOST

"Cock-a-doodle-doo," announces a rooster masked by dense forest in Kokee State Park, an achingly beautiful nature reserve on Kauai's western coast, on a sunny autumn morning. A faint but unmistakable "cock-a-doodle-doo" volleys back, from maybe a kilometre away.

Although the birds are a fixture even in this rich and remote landscape, most of the chickens in the reserve stick near the car parks and picnic areas, where human hand-outs are easy to come by. The park's birds are among the most brazen and comfortable around humans in all of Kauai, and it's hard to enjoy a meal in Kokee's central meadow without attracting the attention of a flock or two. But "give them a chase and they'll disappear down a 300-metre ravine that's so thick with vegetation it's impossible to follow," says Gering. "That's not something I think about barnyard chickens as being able to do."

A park website discourages visitors from feeding the chickens, in the hope of reducing their numbers and their dependence on humans. This interest in 'rewilding' the feral chickens is probably motivated by the desire to reduce their numbers through methods other than culling. It may well be a matter of time before feral animals fully shed their yokes and evolve into creatures less dependent on humans — but perhaps it will never happen. "The environmental niches that feralized animals are exploiting are very different and bear a human stamp that wasn't there when their ancestors developed," says Zeder. "Why should anyone expect a feralized animal to go out and become the noble savage animal again?"

Wright, however, thinks there is a possibility that if the chickens in Kokee are left alone for long enough, they may well become not a facsimile of their red-junglefowl ancestors, but some other kind of creature just as deserving of being called wild. Whatever wild means. ■

Ewen Callaway writes for Nature from London.

1. Gering, E., Johnsson, M., Willis, P., Getty, T. & Wright, D. *Mol. Ecol.* **24**, 2112–2124 (2015).
2. Feulner, P. G. D. *et al. Mol. Ecol.* **22**, 4210–4221 (2013).
3. Anderson, T. M. *et al. Science* **323**, 1339–1343 (2009).
4. Maselli, V. *et al. Evol. Biol.* **41**, 229–239 (2014).
5. Der Sarkissian, C. *et al. Curr. Biol.* **25**, 2577–2583 (2015).



Instruments on a 50-metre-tall tower in western Virginia will monitor the metabolism of the forest by measuring carbon dioxide and water vapour.

ECOLOGY AIMS HIGH

The United States is sinking nearly half-a-billion dollars into a grand ecological observatory, but the project has been dogged by budget overruns and delays.

BY CHRIS CESARE

Overgrown shrubs thwack the sides of a pick-up truck as it bounces along a dirt road through a forest in western Virginia. On the drive, ecologist Ty Lindberg calls out the names of the invasive species crowding either side of the path. There is mile-a-minute weed, which spreads with alarming speed. A flowering annual called Asian stiltgrass carpets the ground and stifles native plants. And a particularly prickly species of rose tears at the clothes of anybody who ventures too close. “My field techs don’t enjoy that one,” Lindberg says.

Farther along, he stops the car at a break in the brush and picks his way through the undergrowth towards a set of plastic and aluminium stakes poking out of the ground. They mark

out a 40-by-40-metre plot, one of dozens scattered throughout 1,300 hectares of forest and pasture near the town of Front Royal. From April to October, field technicians spend their days cataloguing the location, diameter and height of nearly every tree in the plot, collecting fallen leaves out of a trap and archiving pressings from invasive plants. Their main goal is to measure the ecosystem’s metabolism, especially how much biomass it generates each year.

At other plots, technicians trap rodents and draw blood samples to test for diseases, including those that could spread to humans. The staff collects and stores ticks and beetles, and takes soil samples to study the bacteria underfoot. Higher up in the hills, a 50-metre-tall metal tower juts above the trees, loaded with long booms holding sensors that monitor air temperature, wind

speed and solar radiation at multiple altitudes. When the tower has its final instrument package installed in 2016, it will watch the forest breathe by monitoring how carbon dioxide and water vapour concentrations rise and fall.

This site is one of more than 80 planned for the National Ecological Observatory Network (NEON), a US\$434-million project to build a biological observatory that spans the United States. Its goals are grand. If all goes well, it will document the effects that climate change and land use have on ecosystems and provide scientists with a nearly real-time measure of the country’s ecological vital signs. Many of the sites will operate for three decades, whereas others will be packed up and relocated periodically in response to environmental changes. And the data collected will be freely

CHRIS MADDALONI/NATURE

SOURCE: NEON

available to all through an online portal.

Lindberg, who manages three NEON sites in Virginia and Maryland, says that the long-term record generated by NEON could transform ecology by helping scientists to answer questions ranging from how invasive species are altering the landscape to how quickly infectious diseases are spreading through ecosystems. The network, he says, “is really an instrument”. Ecologists call it their first foray into big science — a massive project that rivals the scale of big-budget physics facilities such as particle colliders or telescopes (see ‘Sentry posts’).

But ecologists have not had an easy journey into the world of big science. During its five-year construction phase, NEON has encountered a series of high-profile problems that have raised concerns about the programme, which is funded entirely by the US National Science Foundation (NSF). In June 2015, the network came under fire from the NSF and Congress after NEON, Inc. — the non-profit organization that manages the project — reported that it was running \$80 million over budget. Amid revelations that the company had spent federal money on parties, Congress levied charges of mismanagement and convened hearings with officials from NEON and the NSF. Events came to a climax in December, when the NSF decided to take NEON, Inc. off the project, citing a lack of confidence in the company after years of delays and questions about accounting irregularities.

The agency will now seek another operator to complete construction and take over the project’s management. One of the toughest tasks will be winning the support of ecologists; some researchers felt alienated during the project’s planning phase and have been critical of the way the observatory network is turning out.

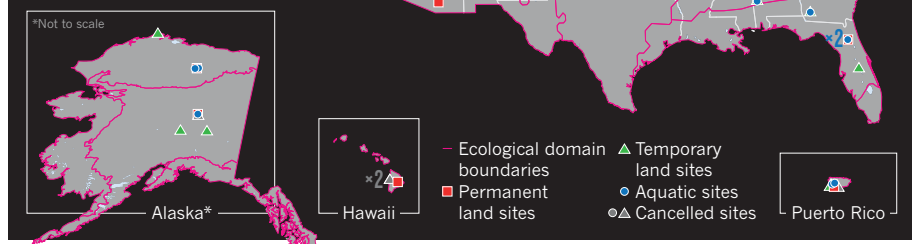
Still, many ecologists are eager to get their hands on NEON’s data and are already thinking about how to incorporate it into their studies. Ultimately, the science that they produce will determine whether the project succeeds or fails. “You build out NEON and in 30 years you’re going to have unprecedented data on how ecosystems are changing,” says Peter Groffman, an ecosystem ecologist at the City University of New York. “It’s very exciting and very much the next logical evolution of long-term studies.”

BIG BIOLOGY

The debates about NEON reach back to its early evolution, when it took shape in a manner very different from a major physics project. Scott Collins, an ecologist at the University of New Mexico in Albuquerque, was the first NSF programme director for NEON back in 2000. Collins says that the idea for a large ecological observatory sprang from NSF staff who were seeking ways for biologists to get a slice of the agency’s big-science money: the Major Research Equipment and Facilities Construction budget. “That put us on a very different footing from the start because this was not

SENTRY POSTS

Nearly half of the 81 sites that make up the National Ecological Observatory Network (NEON) are permanent; the others will be moved periodically. The sites are spread across 20 ecological domains. Fourteen extra stations were cut because of budget problems.



“I THINK IT’S GOOD FOR SCIENTIFIC COMMUNITIES TO DREAM BIG.”

something that the community and vocal ecologists had wanted,” Collins says.

Although researchers did not dream up the project, they quickly embraced the idea and took the lead in moulding NEON’s design during workshops. At these meetings, attendees were encouraged to dream big, says Katherine Gross, the director of Michigan State University’s W. K. Kellogg Biological Station in Hickory Corners and the current chair of the NSF’s biological-sciences advisory committee.

During six workshops between 2000 and 2002, ecologists developed a plan for a flexible network of observatories and experimental centres spread across at least ten sites. But scientists disagreed over whether the NSF should specify research themes for each site or allow ecologists to choose their own focus.

In its 2002 budget request to Congress, the NSF asked for \$12 million to develop and build two prototype NEON sites, largely based on the reports from the workshops. But Congress denied the request, citing a lack of information about the project and an insufficient estimate of its costs.

At the time, the best model that US ecologists had for NEON was the Long Term Ecological Research network, a group of investigators in the United States who study the ecology of a particular spot for five or more years with sets of measurements specialized to each site. The leaders of NEON, however, eventually settled on a one-size-fits-all approach, with standard protocols and instruments that could be deployed across the entire network to study pressing issues, including changes to biodiversity and climate change. And instead of having principal investigators propose individual observatories,

an expert panel recommended that the NSF develop NEON as a nationwide network managed by one entity.

Encouraged by Congress to continue refining its idea of NEON, the NSF issued a call for proposals in early 2004. A \$6-million grant to design NEON went to several members of the American Institute of Biological Sciences in Reston, Virginia, a non-profit organization that had been involved in the project’s earlier planning. A little more than a year later, in December 2005, the lead designers created NEON, Inc.

Over the next several years, the structure of NEON took shape. The network split the country into 20 domains, each with several sites outfitted with instruments and collection protocols.

FEELING IGNORED

But when it came to choosing where to build sites and how best to make measurements, some ecologists objected to the choices and felt that their expertise had been ignored. Gene Kelly, a soil scientist at Colorado State University in Fort Collins — and now the interim chief executive of NEON, Inc. — says that the emphasis on measuring the same quantities everywhere meant that NEON had to sacrifice having the optimal protocol for every spot. “The only way to really handle it was to standardize it, but in doing that you lose a little,” he says.

Kelly says that many ecologists, including him, stopped following the progress of the project closely after these decisions were made, partly because NEON, Inc. stopped asking for input.

Despite the loss of some engagement from the ecological community, the NSF approved

NEON's final design in 2009, and Congress authorized the money for construction in July 2011. The network would spread 17,000 sensors measuring hundreds of variables — from soil moisture to stream pH — across nearly 100 sites. And at each site, technicians would collect a suite of biological samples, including genomic data from many organisms and whole specimens of insects and small mammals. The result would be a standard set of ecological data that would allow scientists to compare and watch for changes in ecosystems and to produce ecological forecasts.

Concerns about the company's accounting and the NSF's oversight cropped up almost immediately after construction began in 2012. A review that year found that NEON's books were a mess: auditors questioned more than one-third of the total construction cost — \$154 million — and determined that NEON, Inc. did not provide enough information to support its proposed budget.

Later audits and investigations unearthed questionable spending by NEON management, including \$25,000 for a party and \$3,000 for T-shirts. Also, the company moved to a new office and paid nearly \$500,000 for time left on the old lease. After the audits, the NSF's inspector-general urged the agency to keep a closer eye on the project.

Despite the accounting problems, the NSF and NEON, Inc. forged ahead with construction — and ran straight into delays. Some could have been predicted, such as the difficulty of obtaining permits to build observation towers in cities. Others were simply bad luck. At the site in western Virginia, a tree fell over and destroyed a collection of atmospheric instruments. A bear damaged fibre-optic data lines running to soil-monitoring instruments near the Virginia site's tower. And concerns about the health of a pregnant cheetah at a nearby conservation facility forced NEON, Inc. to abandon plans to use a helicopter to hoist the topmost sections of the instrument tower into place. Instead, construction staff raised the final sections by hand.

The delays put NEON behind schedule and over budget. In June 2015, the company told the NSF that it would take an extra \$80 million on top of the \$434-million budget to complete construction.

With Congress already concerned about the NSF's stewardship, the agency demanded that the project be downsized to stay within its budget. It told NEON, Inc. to cull the number of sites from 95 to 81, cancel construction of its stream experiments and give up some of its embedded sensors. NEON, Inc. then fired its chief executive last September and appointed Kelly to serve as an interim. But the company sealed its fate in December when it submitted an updated budget that again had extra costs and delays. The NSF decided to look for a new company to manage the project.

Whoever takes over will step into a difficult



Clockwise from top: a camera monitors stream depth, a tag marks invasive species and a marker delineates plots at a site in western Virginia.

role, as many ecologists remain disconnected from the project. Yet there are still big hopes for NEON in the research community. "I think it's good for scientific communities to dream big and say, 'OK this will be our unifying project,'" says Ash Ballantyne, a bioclimatologist at the University of Montana in Missoula. "It's analogous to our LHC."

GLOBAL REACH

Interest is growing as money starts to flow towards individual researchers. In August, the NSF awarded \$4.8 million in grants to investigators and workshop organizers who are interested in using NEON data. Ballantyne received \$300,000 to study the effects of drought, fire and insect infestations on the carbon cycle. He plans to investigate how drought or other disturbances predispose trees to a beetle outbreak or fires.

Jim Clark, an ecologist and statistician at Duke University in Durham, North Carolina, won a grant to build more-sophisticated ecological models. "We've always modelled on a species-by-species basis," Clark says. "If there's 100 species, someone has fitted 100 different models and just added them together." This ignores the interactions between species, he says, and NEON data on species abundance could help to fit and train joint models for how species respond to ecosystem changes collectively.

Frank Davis, an environmental scientist at

the University of California, Santa Barbara, says that he plans to use NEON's airborne observations to study tree cover at various scales, from a few centimetres to several kilometres. Many ecologists are not accustomed to thinking at the large scale that NEON covers, Davis says. "Ultimately, I think NEON will be ready for ecologists," he says. "But will ecologists be ready for NEON?"

Some are gaining valuable experience thinking at global scales by running their own distributed networks. The Global Lakes Ecological Observatory Network began in 2005 and ties together groups around the world that monitor human effects on lake ecosystems. The Nutrient Network, or NutNet, links researchers on six continents who perform a standard set of experiments looking at how plant production in grasslands is limited by phosphorus and nitrogen — two by-products of fossil-fuel combustion. Other networks are springing up to study plant populations and drought.

These projects are smaller in scope than NEON, which gives researchers more control over the work. With only a handful of voices deciding how to conduct experiments or take data, projects such as NutNet can maintain a tight focus on the science. "It's very hard for NEON to do this because the entire ecological community has a say," says Elizabeth Borer, an ecologist at the University of Minnesota in St Paul and a member of NutNet.

Ecologists are still struggling to learn how to manage large projects, says Nikki Thurgate, an ecologist at the University of Adelaide in Australia and leader of international engagement for the Terrestrial Ecological Research Network — a smaller, Australian cousin of NEON. But if ecology is to forecast the problems that arise from climate change and loss of biodiversity around the world, it will need the data from large-scale networks. And one of the challenges is to keep the community engaged and informed while they wait for a grand scientific instrument to be built. "You can't pop up continent-wide environmental monitoring and have data in a couple of years," Thurgate says. "It's just not that simple."

NEON's early struggles may fade when data start to arrive in the next few years from sites such as the one that Lindberg manages in western Virginia. On a cold day late last year, Lindberg — who still has his job for the time being — stood below the nearly finished observation tower rising high above the surrounding forest. In a nearby shed, dozens of boxes held sensors and electronics to be installed on the tower. Despite the work that remains here and at other sites around the country, Lindberg still thinks that the project can be successful — as long as researchers embrace it. "It's a scaffolding," he says. "But this thing doesn't work unless scientists use it." ■

Chris Cesare is an intern with *Nature* in Washington DC.

COMMENT

GENOMICS African American ancestry-seekers deserve good data **p.279**



INTERDISCIPLINARITY A manual for forging great collaborations **p.280**

POLICY Experts raise alarm about peer review of Japanese whaling **p.283**

OBITUARY Nobel winner Alfred Goodman Gilman, discoverer of G-proteins **p.284**

PAUL D. STEWART/NATUREPL.COM



Studying the feet of a variety of gecko species would help scientists to design a richer variety of dry adhesives.

Bring biologists into biomimetics

Engineers, chemists and others taking inspiration from biological systems for human applications must team up with biologists, writes **Emilie Snell-Rood**.

In the late 1940s, Swiss engineer George de Mestral was inspired to invent Velcro after picking burrs from the fur of his dog. Five decades later, Japanese engineer Eiji Nakatsu, who had a passion for bird watching, designed a high-speed train with an aerodynamic front shaped like a kingfisher's beak.

Over the past decade, there has been a surge of interest in such bioinspiration and biomimicry — the imitation of biological traits or systems in applications ranging

from architectural design and materials to robotics and engineered tissues^{1,2} (see 'Trends in biomimetics'). The fervour is likely to be amplified by the emergence of the cheap and easy-to-use genome-editing tool CRISPR-Cas9, which has dramatically opened up the possibilities of altering organisms or their products.

Biomimicry successes in recent years include the mimicking of aquaporin molecules — proteins that form pores in the membranes of cells — for desalinization;

the production of spider silk (which has a unique combination of strength and elasticity) from genetically engineered silkworms³; and the development of catalysts akin to components used in photosynthesis to improve the efficiency of solar power⁴.

TAP EXPERTISE

Bioinspiration is poised to play a major part in efforts to solve problems relating to health, energy efficiency and food security. To realize this promise, researchers ►

► exploring such approaches need to take much greater advantage of the knowledge and expertise of biologists — be they ecologists, microbiologists or specialists in evolution, organisms, cells or molecules.

The field of biomimicry mostly involves chemists, engineers and materials scientists. Fewer than 8% of the nearly 300 studies on biomimetics published in the past 3 months and indexed in the Thomson Reuters Web of Science had an author working in a biology department — a crude proxy for ‘a biologist’. And in most papers on biomimetics, the relevant biodiversity gets short shrift. In more than 80% of papers published on biomimetics in the past year, for instance, researchers consider only one species or refer to a biological element such as a ‘cell’ or ‘enzyme’ in only a generalized way. Moreover, across studies that explore diverse processes and systems, many of the same players come up: geckos, spiders and butterflies.

With around 1.5 million described species, and probably some 9 million eukaryotic species in existence, researchers pursuing biomimetic approaches have barely scratched the surface of biological inspiration. Biologists from all sorts of disciplines have an extraordinary store of knowledge that could guide forays into a much richer diversity of natural systems. Such knowledge could also help to steer experimental approaches.

On the one hand, an understanding of a particular organism’s biology can inspire specific applications. Naked mole rats, for instance, which can live up to about 30 years — an order of magnitude longer than related rodents — may have adaptations that are relevant to the treatment of age-related diseases such as cancer. Indeed, experiments have revealed that an extracellular substance makes the cells of naked mole rats remarkably tumour-resistant⁵.

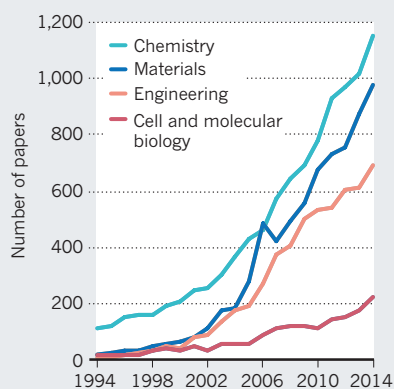
On the other hand, biologists can match a problem that a researcher may be trying to tackle to a specific species, environment or suite of evolutionary conditions. For example, an understanding of the general conditions that favour the evolution of altruism — such as a high probability of repeated encounters with recognizable individuals — may offer insights into how to design architectures and cityscapes that promote cooperation between people.

Evolutionary biologists, in particular, could help biomimicry researchers to use entire lineages of species to test general principles about how form and function relate to the environment. For instance, comparative studies of spiders indicate that different structural properties of silks have evolved as species have adapted to different environments⁶.

Comparative studies may also suggest ways around the limitations of copying a particular trait. Traits that have arisen

TRENDS IN BIOMIMETICS

A search of the more than 25,000 papers in biomimicry shows the rising interest in the field over the past decade, but studies are mainly restricted to the physical sciences.



Data obtained by searching the Web of Science Core Collection with the term ‘biomim*’ or ‘bioinspir*’.

through evolutionary processes, which depend on myriad factors such as the availability of genetic variation, are rarely perfect from an engineering perspective. A robotics engineer would not include a blind spot in a visual system, for instance. So researchers should draw inspiration from multiple systems; the camera-type eyes (like those of humans) that have evolved independently in octopuses do not have a blind spot. Many of the traits of interest in biomimetics — such as the dry adhesion of gecko toe pads or the iridescent coloration of butterfly wings⁷ — have evolved many times independently. This offers researchers different ways to produce a trait of interest.

Finally, biologists can help biomimicry researchers to explore processes that could make applications more feasible. Reproducing the dry adhesion of a gecko toe pad or the strength of a composite material such as abalone shell is difficult because of the complex nature of the structures involved². To produce such materials economically at scale, some in biomimetics have turned to investigating how structures are built as well as the structures themselves. For example, photonics researchers, interested in manipulating light, have begun trying to understand the developmental processes that produce butterfly-wing scales⁸. And directed evolution is being used in the laboratory to produce new biocatalysts from pre-existing enzymes⁹.

EMBRACE TEAMWORK

Various funding sources are increasingly supporting research that encompasses blue-sky biological research and more applied fields. Translational-research centres, such as the Engineering Research Centers programme funded by the US National Science Foundation, are being established to help to tie basic research to the development of

applications in industry¹⁰.

Four further steps would encourage more crossover between biology and biomimetics.

First, changes to education could encourage budding biomimicry researchers to embrace more biology. Currently, undergraduates who end up pursuing work in biomimicry are trained in the ‘application field’ of their choice — be it engineering, architecture or chemistry. Such students should be required to take courses in the area of biology that is likely to be most relevant to them: perhaps molecular biology for those pursuing chemistry, or organismal or evolutionary biology for those pursuing materials science or architecture.

Second, we need textbooks that distil the essential principles of specific branches of biology for biomimetic approaches; existing texts focus more on the applications.

Third, matchmaking initiatives are needed to promote collaboration. Existing conferences on biomimetics — such as the International Conference on Bioinspired and Biobased Chemistry and Materials, or the International Conference on Bio-inspired Systems and Signal Processing — could offer symposia or incentives to attract biologists. Special events in the style of speed dating could allow biologists and those interested in bioinspired applications to partner up based on their probable ability to solve a problem together. Online matchmaking sites might serve a similar purpose.

Last, companies could bring in broadly trained evolutionary and other biologists, for instance through ‘biologist-in-residence’ programmes. Such collaboration would fast-track robust solutions, by enabling teams to explore innovations from across the tree of life. ■

Emilie Snell-Rood is an assistant professor of ecology, evolution and behaviour at the University of Minnesota, Saint Paul, Minnesota, USA. She has designed courses in biomimetics and is developing butterflies as a system for biomimetic approaches to health.

e-mail: emilies@umn.edu

1. Bhushan, B. *Phil. Trans. R. Soc. A* **367**, 1445–1486 (2009).
2. Wegst, U. G. K. *et al. Nature Mater.* **14**, 23–36 (2015).
3. Teulé, F. *et al. Proc. Natl Acad. Sci. USA* **109**, 923–928 (2012).
4. Kurz, P. in *Solar Energy for Fuels* (eds Tüysüz, H. & Chan, C. K.) 49–72 (2016).
5. Tian, X. *et al. Nature* **499**, 346–349 (2013).
6. Swanson, B. O. *et al. Evolution* **60**, 2539–2551 (2006).
7. Vukusic, P. & Sambles, J. R. *Nature* **424**, 852–855 (2003).
8. Saranathan, V. *et al. Proc. Natl Acad. Sci. USA* **107**, 11676–11681 (2010).
9. Otten, L. G. & Quax, W. J. *Biomol. Eng.* **22**, 1–9 (2005).
10. Bozeman, B. & Boardman, C. J. *Technol. Transfer* **29**, 365–375 (2004).



US slaves' families were deliberately broken up, so it is hard for descendants to trace their roots.

GENOMICS

DNA and diasporas

Fatimah L. C. Jackson weighs up a study on the cultural politics of genetic testing among African Americans.

Genetic analyses, including research on the origins of Europeans, have been invaluable in clarifying disputed or unresolved aspects of population history. Yet studies of the genetics of African diasporas, including those in the Americas, are in their infancy. This is largely down to a lack of interest among majority-population researchers, limited historical knowledge of the groups' origins and dispersions, and a paucity of meaningful interdisciplinary collaborations between researchers exploring population substructure in the various African diasporas.

The Social Life of DNA by sociologist Alondra Nelson addresses the cultural politics associated with the nascent genetic insights into African American history. She competently explores many of the sociological ramifications of relying on commercially available genetic data that promise to fill in the genealogical gaps in African American lineages and provide the potential to heal the psychological wounds of lost ancestral knowledge. However, two fundamental elements are missing from the book. One is a robust

critique of the quality of the genetic findings discussed. The other is a thorough, up-to-date examination of the medical mistrust that is likely to grow among this vulnerable population once the fallacies and limitations of the current genetic-ancestry paradigm are more broadly recognized. The main problem is that 'precision medicine' is tailored on the basis of ancestry. So, because the ancestry data for people of African descent is incomplete and sometimes incorrect, the potential for providing precision medicine to these individuals is extremely limited.

Interest in the genetic history of African diasporas was spurred by complaints from African Americans that they had initially been excluded from both the Human

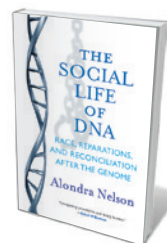
Genome Diversity Project and the Human Genome Project. These complaints led a group of African American social and life scientists to issue the Manifesto on Genomic Studies among African Americans in 1994. Soon, more federal funding became available to study health disparities in this group, primarily through the emergence in 2010 of the National Institute on Minority Health and Health Disparities (NIMHD), part of the US National Institutes of Health in Bethesda, Maryland. These infrastructural changes have resulted in increased funding for, and hence recognition of, the contributions of African American genetic and, later, genomic studies to differences in health outcomes. This body of research indicates that there are high levels of genomic heterogeneity within this population, so studies that take a more detailed account of an individual's geographic and historical background are required.

Improvements in technology and sampling have allowed researchers to detect 'ancestry-informative markers,' such as the significant differences in population frequency of particular blood group alleles. The question of whether we can meaningfully set parameters on African American diversity took on a genetic focus with the rise in direct-to-consumer genetic testing and interpretations based on limited uniparental markers.

By 2004 it became clear that some ancestral verification could be achieved for African Americans through the study of mitochondrial DNA (mtDNA) from New York City's African Burial Ground National Monument. An estimated 15,000 people were buried here during the seventeenth and eighteenth centuries; more than 400 bodies have been recovered and evaluated. The database of mtDNA, Y-chromosome and autosomal markers has grown, but remains incomplete and inadequate for comprehensive 'root-seeking': moreover, the contemporary surrogates for a meaningful number of African American ancestral populations are not yet included.

Nelson does not mention it, but the Cobb Research Laboratory (which I direct) at Howard University in Washington DC has begun to augment that database by extracting DNA from skeletal, dental and bioarchaeological samples going back over 400 years in African American history. Although there have been significant improvements in our understanding of the identity and distribution of several African diasporic populations, we are far from understanding the source populations of Africans in the Americas, or the environmental factors that influence gene-expression patterns of African Americans today.

The peculiarities of US slavery and the cruelties of Jim Crow racism — the enforcement of state and local laws that allowed segregation into the 1960s — after emancipation systematically destroyed not only families but the foundations of a robust knowledge of



The Social Life of DNA: Race, Reparations, and Reconciliation After the Genome
ALONDRA NELSON
Beacon: 2016.

► ancestral identity among African Americans. In his 1845 *Narrative of the Life of Frederick Douglass*, the eponymous, formerly enslaved abolitionist and statesman exemplified the quest for that identity, writing: “A want of information concerning my own was a source of unhappiness to me even during childhood.”

The inability of African Americans to access their original names, language, tribe and clan affiliations or regional origins in Africa has left wounds in the psyches of many, and has made ancestry reconstruction a high priority. The intentional destruction of families during slavery contributes to the damage, and to the complexity of the search. Finding relatives separated by enslavement was usually a person’s first task after emancipation; often full reunification was impossible.

In the United States, identity is disproportionately linked to phenotype, and in ethnic-minority groups is associated with levels of self-confidence, performance, and overall positive life outcomes and behaviours. Commercial purveyors of ‘direct-to-consumer’ genetic-ancestry testing have identified a vulnerable and accessible market in many African Americans. Unfortunately, the results fall



Researchers at Howard University extract DNA from samples.

decidedly short of the marketing hype. Many scientists deem these tests and their interpreted results ‘recreational genetics’; they are not peer-reviewed, and work with unknown reference databases. Companies have sequestered information from public databases to build their own proprietary databases, which are inaccessible to scientific scrutiny.

Today, a perfect storm rages at the nexus of commercial opportunism (targeting, moreover, a socio-economically disadvantaged

group), biomedical and technological inadequacy, historical naivety and the scars of the institutionalized loss of ancestral ties. How will inaccurate and inadequate genetic-ancestry information be reconciled with the promise of personalized, precision medicine when the latter depends on accurate and comprehensive genetic ancestry data? Perhaps this will be the topic of another sociological study.

The Social Life of DNA is a lucid, if insufficiently critical, chronicle of the sociocultural maturation of a population as its individuals use new resources to recognize and develop an expanded set of identities. The book will date quickly because this process, and the science undergirding it, is dynamic. Still, it is an admirable first

effort to explore the initial implications of genetic root-seeking among people of African descent, with much room left for adding a necessary, in-depth biological perspective on the topic. ■

Fatimah L. C. Jackson is a professor of biology at Howard University in Washington DC and director of its W. Montague Cobb Research Laboratory. e-mail: fatimah.jackson@howard.edu

MULTIDISCIPLINARITY

Collaborative to the core

Rick Ryland applauds a study on the lean new model for research teamwork.

The management of research processes and systems has evolved significantly over the past few decades. There has been, for instance, an erosion of the distinction between ‘pure’ and ‘applied’ work, conceptually and in practice. And there has been a rise in collaborative relationships, both between universities and between universities and other agencies such as businesses, public-sector institutions and charities. This process is becoming increasingly international.

In *The New ABCs of Research*, Ben Shneiderman examines the causes of these changes. Some will be familiar. Interdisciplinary work encourages diverse partnerships, and electronic communications enable rapid interaction and dissemination. Research teams are replacing what Shneiderman calls the myth of the lone researcher. The pursuit of both commercial advantage and public benefit are stimulating collaboration; data and resources are being pooled; and governments are supporting international research cooperation for diplomacy and trade. Shneiderman,

a specialist in information technology and information systems, delivers a shrewd and timely account of these developments.

He champions collaboration, teamwork and practice-based research. The book intends to serve as a guidebook for “students and junior researchers” and as a “manifesto for change, particularly for senior researchers and policymakers”. There is some tension between these aims. But his argument is coherent: senior scientists and policymakers do need to adapt the present to assist the future, and the generation now being trained, or starting their careers, will carry the torch. Shneiderman has stimulating thoughts on



The New ABCs of Research: Achieving Breakthrough Collaborations
BEN SHNEIDERMAN
Oxford University Press: 2016.

project-based teaching at the undergraduate level and advice for early-career researchers on winning grants, publication and effective writing. All this is hands-on and helpful, and one wishes it had been said more loudly in the past. But his proposals for policymakers do not extend much beyond exhortation.

The core principles of *The New ABCs of Research* are delivered in two recurrent acronyms. The ABC of the title stands for “applied and basic combined”, whereas SED is “science, engineering, and design”. The latter, Shneiderman argues, can deliver the former (which also stands for, in the book’s subtitle, “Achieving Breakthrough Collaborations”, Shneiderman’s stated ambition).

Of the three SED elements, Shneiderman is unadventurous on science and engineering; on design he is more stimulating. Drawing on contemporary theories of “design thinking”, he notes that design is not a matter of efficiency or refinement, but rather a “human-centered process” essential for the discovery of meaning and function. He sees it as imbued

with the empathy necessary for identifying human needs, “which are then distilled into an actionable clarification of the problem”. Shneiderman is ambitious, envisioning collaboration between designers, scientists and engineers to tackle globally important issues.

The book is strongest in the areas in which Shneiderman is most enthused, such as the advantages of teamwork or the concept of prototyping as research in practice, for instance using 3D modelling and computer-assisted virtual design tools. His discussion of the challenges and potential of big data and open access as ways of sharing research findings and approaches to social issues is informed and thoughtful. There is a refreshing pragmatism about his attention to matters such as the business of making a research career. Shneiderman is good on career psychology, and on the scarcity of women at professorial levels in particular. He is especially good on the human dynamics of collaboration. Alongside sensible advice on the dos and don'ts of partnerships and the optimal size and mix of skills in groups, he pays attention to matters such as leadership, status management, brokerage, goal-setting and communication among different personality types.

But *The New ABCs of Research* can be over-schematic and repetitive. Its acronymic principles pop up like alerts on a smartphone. Clichéd metaphors (for example seeds, root and flowers) stand in for descriptions of the dividends of interdisciplinary work or painstaking policy recommendations. Zealotry sometimes erodes the book's gravitas.

Shneiderman provides several case studies, including sketches of ABC-effective research organizations, such as the German Fraunhofer Institutes, or Bell Labs in the United States. Most interestingly, he discusses individuals including Nathan Eagle, founder of Jana, a company that specializes in deploying mobile technologies as local solutions to issues in the developing world, for example to encourage blood donation. But these studies tend to serve not as moments for exploration, but as campaigning examples of ABC in practice. That over-insistence is a touch reminiscent of pop business-studies books for under-confident entrepreneurs.

Like several books of its ilk, I wished it was both longer to substantiate the case, and shorter to get past the too-readily recognizable. But overall, *The New ABCs of Research* deserves a readership for the boldness of its ambition and the promise of its ideas. They deserve debate and implementation. ■

Rick Rylance is director of the Institute of English Studies at the University of London's School of Advanced Study. He was until recently chief executive of the Arts and Humanities Research Council and Chair of Research Councils UK.
e-mail: rick.rylance@sas.ac.uk

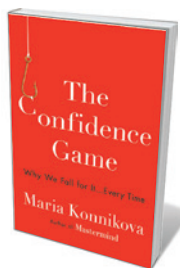
Books in brief



The Shock of the Anthropocene: The Earth, History and Us

Christophe Bonneuil and Jean-Baptiste Fressoz (translated by David Fernbach) VERSO (2016)

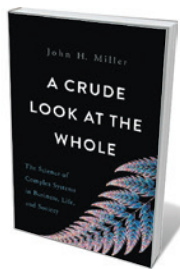
This bold, brilliantly argued history of the Anthropocene epoch is a corrective to cosy thinking about humanity's grave disruptions to Earth systems. Christophe Bonneuil and Jean-Baptiste Fressoz draw on climate science, economics and technological history to reveal how, starting in eighteenth-century France, imperial narratives that saw people and planet as a “totality to be governed” laid the conceptual basis for the crisis. They call for a “new environmental humanities”, and a shift away from market-based approaches that feed the beast.



The Confidence Game: Why We Fall for It... Every Time

Maria Konnikova VIKING (2016)

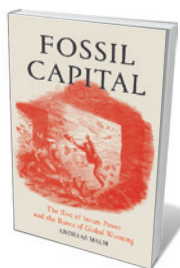
Following her nifty how-to on honing cognitive ability, *Mastermind* (Viking, 2013; see *Nature* **492**, 183; 2012), journalist Maria Konnikova adroitly explicates the surprising psychology behind the confidence game — the modus operandi of charismatic swindlers that thrives in upheavals such as today's technological revolution. She unpacks the con-artist's repertoire of cajolery, illustrating it with case studies (such as art dealer Glafira Rosales's large-scale fraud) and research (including psychologist Paul Ekman's, on lying). A mesmerizing glimpse into the trickster's mind.



A Crude Look at the Whole: The Science of Complex Systems in Business, Life, and Society

John H. Miller BASIC (2016)

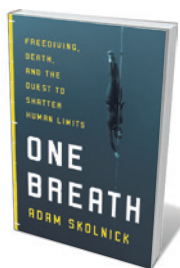
Reductionism offers few insights into complexity in nature. So argues computational analyst John Miller in this succinct, elegant study of systems thinking, the newish science examining basic principles, such as emergence, that govern physics, biology and economics. Miller reveals compelling echoes between apparently unrelated phenomena, such as “hivemind” behaviour in bee colonies and consumers, or responses to local stimuli in how a cone snail patterns its shell and how a market functions.



Fossil Capital: The Rise of Steam Power and the Roots of Global Warming

Andreas Malm VERSO (2016)

The birth of the fossil economy, avers human ecologist Andreas Malm, arrived when steam eclipsed water power in mid-nineteenth-century Britain. Around that, Malm builds a deep, insight-packed history of how society came to be in thrall to the twin engines of combustion and capital. We see, for instance, how at the start, steam was simply more expedient, not more efficient, than hydropower; and how now, decoupling from fossil fuels is stymied when energy companies pull out of investment in renewables on the basis of low returns.



One Breath: Freediving, Death, and the Quest to Shatter Human Limits

Adam Skolnick CROWN ARCHETYPE (2016)

Freediving (making deep dives on one breath) has been having a moment since James Nestor's *Deep* surfaced (Houghton Mifflin Harcourt, 2014; see *Nature* **510**, 339; 2014). Here, Adam Skolnick interlaces the science of the sport with the story of US freediver Nicholas Mevoli, who died in competition in 2013. Pulmonary haemorrhages contributed to his death, Skolnick shows, pointing to a need for more research on this radical self-experimentation. *Barbara Kiser*

Correspondence

Japan's whaling is unscientific

Japanese whalers are back in the Southern Ocean, aiming to kill 333 minke whales — ostensibly for the purposes of scientific research — under special permits issued by their government. In our view, the science behind Japan's whaling activity has not passed a reasonable standard of peer review.

We are members of the International Whaling Commission's Scientific Committee (IWC-SC), plus one independent expert witness (M. Mangel) whose evidence contributed to the March 2014 negative ruling on Japan's JARPA II whaling permit by the International Court of Justice (ICJ); see A. S. Brierley *Nature* **520**, 157; 2015). In 2015, Japan submitted a new whaling proposal (NEWREP-A). The IWC-SC coordinated two rounds of review, including one by an independent expert panel that concluded that lethal sampling had not been justified. Numerous IWC-SC members recommended exploration of widely used non-lethal alternatives (see, for example, A. M. Polanowski *et al. Mol. Ecol. Resour.* **14**, 976–987; 2014) before killing is resumed.

Japan claims to have “sincerely taken into account” the IWC-SC's opinion, but, as on previous occasions, has failed to alter its plans in any meaningful way and is proceeding to kill whales under a self-determined quota. In October 2015, Japan also rejected the jurisdiction of the ICJ on this issue.

We believe that further discussion of special-permit whaling at IWC-SC under the present procedure — in which the opinion of proposers is afforded equal weight to that of referees — is a waste of time. The IWC urgently needs to develop a process of scientific review that results in clear decisions that can be respected by all.

Andrew S. Brierley* *University of St Andrews, UK.*

Phillip J. Clapham* *Alaska Fisheries*

Science Center, Seattle, USA.

asb4@st-andrews.ac.uk

**Supported by 30 signatories (listed at go.nature.com/tfb6hp).*

No myth: benefits of breast screening

Your attempt to debunk the “myth” that cancer screening saves lives is ill-conceived for breast cancer (*Nature* **528**, 322–325; 2015). Women should be aware of the benefits and potential harms of breast screening, including that it can reduce the chance of death from breast cancer.

You cite one randomized screening trial that showed no reduction in deaths due to breast cancer. However, an independent meta-analysis of 11 trials, including the one you cite, demonstrates an overall 20% reduction in mortality (M. G. Marmot *et al. Lancet* **380**, 1778–1786; 2012).

Further analysis reveals that only those trials that reduced the incidence of advanced disease (stage II or higher) produced lower mortality (L. Tabár *et al. Breast J.* **21**, 13–20; 2015). This correlation supports the idea that reducing the incidence of advanced breast cancer by using other screening methods might reasonably be expected to reduce mortality. For example, combining mammography with magnetic resonance imaging reduces the chance that women who have a high breast-cancer risk will be diagnosed with late-stage disease (E. Warner *et al. J. Clin. Oncol.* **19**, 3524–3531; 2001).

Screening is also beneficial because early detection reduces the need for mastectomy and chemotherapy. As treatments become more effective, the mortality benefit from screening might be anticipated to decline, but the other benefits will remain.

Jeremy M. Berg, Wendie A. Berg *University of Pittsburgh, Pennsylvania, USA.*

jberg@pitt.edu

No myth: population rise unsustainable

You correctly point out that exponential human population growth is a myth (see *Nature* **528**, 322–325; 2015). The rate of growth has been slowing since 1970, largely as a result of education efforts aimed at women and girls, and family planning programmes. Population growth is still a serious concern, however, because human numbers and increases are currently unsustainable.

Ecological footprint analysis, for example, indicates that it would take 1.6 Earths to support the current world population at average resource-consumption levels (go.nature.com/xyohus). This is set to increase, given the expanded rate of consumption to which developing countries aspire. Moreover, we are tearing the web of life by condemning tens of thousands of non-human species to extinction (G. Ceballos *et al. Sci. Adv.* **1**, e1400253; 2015). A positive rate of growth will make the damaging effects of the human population to our planet even worse.

Richard E. White* *Durango, Colorado, USA.*

rwhite@gobrainstorm.net

**On behalf of 4 correspondents (see go.nature.com/kueb3g for full list).*

Language ruling stymies brain gain

Even a Nobel laureate would now be unlikely to be appointed to the science teaching faculty in many of Spain's best universities. This absurd situation is a result of tenure-track positions becoming conditional on a fluency in Spanish and/or the university's regional language. The restriction, intended to promote local languages, shatters any hope of attracting brilliant minds from abroad into our system.

The national call for 2016 pre-doctoral fellowships, Spain's only public scheme for funding PhD candidates, also demands that

applicants have an intermediate diploma in Spanish or in the official regional language.

We do not question the need to protect all of our languages, which are an important part of our culture. But science is a realm in which content should take precedence over form, the ‘what’ over the ‘how’. We should instead be following the lead of other universities in providing language tuition, if necessary, once a post is taken up.

Spain cannot afford further gambits with its science, which is already seriously compromised by the international financial crisis and questionable short-term science policies.

Pau Carazo, Enrique Font *University of Valencia, Spain.*
pau.carazo@uv.es

Water scheme acts as ecological buffer

We disagree with Jon Barnett and colleagues that China's South-to-North Water Diversion (SNWD) project marks a low point for big engineering solutions to the country's water problems (*Nature* **527**, 295–297; 2015). The SNWD conveyed 8.37 billion cubic metres of water (6.20 billion by the eastern route; 2.17 billion by the central route) to millions of people in northern China by the end of 2015. It is also acting as a buffer against imminent ecological disaster and social upheaval.

It will take time to educate the public about water consumption and to implement the reforms to agricultural and industrial production that the authors propose. The ‘sponge city’ programme to harvest rainwater and recycle wastewater is also likely to be costly and disruptive. Meanwhile, the SNWD is providing the material foundations necessary for well coordinated water management.

Han Lin, Saixing Zeng, Hanyang Ma *Shanghai Jiao Tong University, Shanghai, China.*
zengsaixing@sjtu.edu.cn

Alfred Goodman Gilman

(1941–2015)

Pharmacologist who won a Nobel prize for his discovery of G proteins.

Alfred Goodman Gilman discovered heterotrimeric G proteins, which help to usher chemical signals into cells. For this work, which reshaped our understanding of hormone and drug action, he shared the 1994 Nobel Prize in Medicine or Physiology with Martin Rodbell.

Gilman's impact on biomedical research and education extended much further. He edited several editions of the definitive textbook *The Pharmacological Basis of Therapeutics* (or 'Goodman and Gilman'), which has served generations of medical and graduate students. At the University of Texas Southwestern Medical Center in Dallas, he chaired the pharmacology department from 1981 to 2004, became dean in 2004 and provost in 2006. And in 2009, he was appointed first chief scientific officer of the Cancer Prevention and Research Institute of Texas (CPRIT), established to disburse US\$3 billion in state funding. He died on 23 December 2015.

Gilman was born in July 1941 in New Haven, Connecticut, with, in his words, a "scientific silver spoon" in his mouth. His father was the eminent pharmacologist Alfred Gilman Sr, who wrote the aforementioned textbook with his close colleague Louis Goodman. Gilman Sr, in tribute to his friend, gave his son the middle name Goodman. As a child, young Gilman enjoyed trips to his father's labs at Columbia University and the Albert Einstein School of Medicine in New York.

After completing a degree in biochemistry at Yale University in New Haven in 1962, he enrolled in one of the first MD–PhD programmes in the United States, at Case Western Reserve University in Cleveland, Ohio. The programme was run by Nobel laureate Earl Sutherland, the discoverer of cyclic AMP, a key intracellular messenger molecule. Here, Gilman solidified his interest in cell-signalling mechanisms. A postdoc at the US National Institute of Health in Bethesda, Maryland, with Marshall Nirenberg, another Nobel laureate, followed. In 1970, in Nirenberg's lab, Gilman independently developed a sensitive technique for detecting cAMP that was immediately widely adopted.

But it was his discovery of a family of G proteins made up of three different subunits (known as heterotrimeric G proteins) as a junior faculty member at the University of Virginia in Charlottesville, that transformed the field of cell signalling. In



the 1970s, evidence was mounting that the hormone receptors involved in cell signalling were independent entities in the plasma membrane. Gilman and his postdoc Elliott Ross were investigating this problem using membranes from a lymphoma cell line called *cyc⁻*. These cells seemed to lack the key enzyme adenylyl cyclase, which catalyses formation of cAMP. The cells retained the β -adrenergic receptor that binds a class of molecules called catecholamines, which includes the hormone adrenaline. This binding stimulates adenylyl cyclase, increasing the formation of cAMP.

To these *cyc⁻* membranes, Gilman and Ross added an extract of another cell line, mouse L cells, which retained enzyme activity but lacked the hormone receptors. To their delight, this 'reconstitution' system worked and yielded hormone-sensitive adenylyl cyclase activity. However, control experiments indicated that extracts in which the cyclase activity had been inactivated, by heat for example, still led to a fully reconstituted hormone-sensitive enzyme.

Through ingenious experiments they demonstrated that these extracts contained a heat-stable, hitherto unknown regulatory component required for activity of the adenylyl cyclase and that the *cyc⁻* cells actually did contain the catalytic unit of adenylyl cyclase (and hence had been misnamed). They called the new component $G_s\alpha$ — the α subunit of the stimulatory guanine nucleotide regulatory protein (G protein), because it bound guanosine triphosphate (GTP) and conferred both hormone and GTP

sensitivity on the adenylyl cyclase.

Using their reconstitution assay, they purified the protein and determined its DNA sequence. Their discovery, published in a series of papers in 1977–78 (E. M. Ross and A. G. Gilman *J. Biol. Chem.* **252**, 6966–6969; 1977; and E. M. Ross *et al. J. Biol. Chem.* **253**, 6401–6412; 1978) confirmed Rodbell's prediction of the existence of such an intermediate protein linking a cell's hormone receptors to adenylyl cyclase and intracellular signalling. Over the next 15 years the family of heterotrimeric G proteins grew to about 20, and their roles expanded to include almost all physiological processes in species from yeast to mammals. Gilman's subsequent work with Stephen Sprang focused on the detailed characterization by X-ray crystallography of several of the G proteins and the adenylyl cyclases that they regulate.

Gilman's writing and speaking style were distinctive. He was critical not just of others' work but also of his own. His wit could border on the acerbic if he felt that the highest standards were not being met. He was beloved by his trainees, to whom he was fiercely loyal. They remember him as a hands-off mentor who was nonetheless approachable and inspirational, and who instilled in them a sense of rigour and integrity. Gilman displayed this personal integrity when, in 2012, he resigned from the CPRIT over concerns that political and commercial factors were exerting undue influence on its awarding of grants.

Al's offbeat sense of humour occasionally made its way into his published work. For instance, he designated the irreversibly activated state of adenylyl cyclase induced by a certain guanine nucleotide as the 'P state' (for priapic) because "the enzyme's rate of catalysis was persistently elevated".

Reflecting on the role of serendipity in scientific discovery, Al once commented: "The trick is to recognize good luck when it happens, embrace it, and then commit whatever it takes to extract its full value." He certainly extracted the full value of his lucky start as a scientific blueblood. ■

Robert J. Lefkowitz is a Howard Hughes Medical Institute investigator and professor of medicine at Duke University in Durham, North Carolina, USA. His scientific interests overlapped with Al Gilman's, and they were colleagues and friends for more than 40 years. e-mail: lefko001@receptor-biol.duke.edu

DAVID GRESHAM/UTSOUTHWESTERN

Control of proteasomal proteolysis by mTOR

ARISING FROM Y. Zhang *et al.* *Nature* **513**, 440–443 (2014); doi:10.1038/nature13492

Among the best-studied anabolic actions of mTORC1 are enhancing protein synthesis and inhibiting protein degradation by autophagy^{1,2}. When mTORC1 is inactivated during starvation, proteolysis and autophagy increase to provide amino acids for protein synthesis and energy production. However, Zhang *et al.* recently reported that inhibition of mTORC1 activity for 16 h or more causes a reduction in overall proteolysis by decreasing proteasome expression³. However, their methodology to measure the rates of protein degradation and the rates obtained appear questionable. There is a Reply to this Brief Communication Arising by Zhang, Y. *et al.* *Nature* **529**, <http://dx.doi.org/10.1038/nature16473> (2016).

After radiolabelling cell proteins, Zhang *et al.*³ observed no degradation of these proteins for 8 h, even in the presence of rapamycin, but then proteolysis accelerated and was reduced by rapamycin after a 16 h delay. Such data are inconsistent with several well-established findings: (1) that newly synthesized proteins are degraded rapidly as cells eliminate mis-folded and short-lived regulatory proteins; (2) that proteolysis by autophagy increases within minutes after rapamycin addition^{2,4}; and (3) that degradation of pulse-labelled proteins follows exponential kinetics and thus decreases with time^{5,6}.

In such pulse-chase studies, it is essential to minimize the continued incorporation of radioactive precursors into proteins during the chase period. If the amount of intracellular unincorporated radioactivity is significant relative to the amount in proteins, then continued incorporation of label can mask the loss of labelled proteins by proteolysis and yield misleadingly slow degradation rates. This effect could account for the reported lack of observed proteolysis for 8 h (ref. 3). This problem is particularly serious with short pulses and certain precursors. For example, after labelling for 30 min with ³⁵S-Met/Cys in *Tsc2*-null mouse embryonic fibroblasts, the unincorporated radioactivity equalled 92% of the counts in protein, while with ³H-Phe, the unincorporated radioactivity accounted for only 26% (Fig. 1a). In both cases, significant incorporation continues despite chasing with high levels of non-radioactive amino acids. With ³⁵S-Met/Cys, appreciable unincorporated radioactivity remained in the cells for hours, but much less with ³H-Phe (Fig. 1b). To minimize such problems and measure the degradation of long-lived proteins (the bulk of cell proteins), we label with ³H-Phe for 20 h and chase for at least 2 h before measuring degradation (Fig. 1c)⁷. Consequently, intracellular free ³H-Phe is less than 0.2% of that in proteins, and reincorporation is inconsequential.

By avoiding this recycling problem and conducting our assays in complete medium (without the prolonged serum deprivation used by Zhang *et al.*³), we found that: (1) mTOR inhibition with rapamycin or Torin 1 actually enhanced overall proteolysis immediately in various cells tested⁴, including *Tsc2*-null mouse embryonic fibroblasts used by Zhang *et al.*³ (Fig. 1d); (2) with rapamycin treatment, degradation rates remained increased for 28 h (measured by determining the radioactivity remaining in cell proteins with time (Fig. 1d)), in contrast to the delayed reduction reported after 16 h (ref. 3); (3) proteolysis was also found to increase when we measured rates of release of ³H-Phe from proteins into medium^{7–10} at different times (Fig. 1e); and (4) using selective inhibitors to determine which degradation pathway was activated, we found that mTOR inhibition stimulated proteolysis by both proteasomes and autophagy (lysosomes) (Fig. 1e), as we demonstrated elsewhere⁴.

Thus, we were unable to detect any decrease in proteolysis within 24 h of mTORC1 inhibition as reported by Zhang *et al.*³, but instead

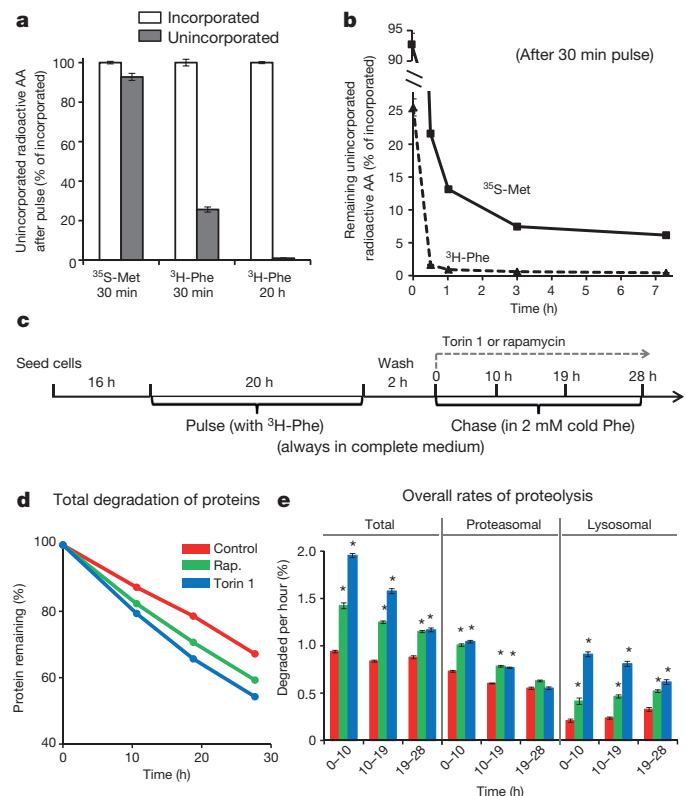


Figure 1 | Pulse-chase methodology using ³H-Phe allows valid measurements of degradation rates and shows that mTOR inhibition causes rapid and sustained increases in protein degradation by both proteasomal and lysosomal pathways in *Tsc2*-null MEFs. **a**, After a 30-min pulse of radioactive amino acids (AA) and repeated washing, large amounts of free (trichloroacetic acid (TCA)-soluble) labelled precursors remain inside cells not incorporated into protein. Cell proteins were labelled by incubation with ³⁵S-Met/Cys for 30 min, with ³H-Phe for 30 min, or with ³H-Phe for 20 h. After three quick washes with chase medium containing 2 mM non-radioactive Met/Cys or Phe, 10% TCA was added to cells to precipitate proteins. TCA-soluble radioactivity (unincorporated precursors) was measured by scintillation counting and expressed relative to the amount of radioactivity incorporated into TCA-precipitable proteins. **b**, After a 30-min pulse, much higher levels of unincorporated ³⁵S-Met/Cys remain in cells for several hours compared to ³H-Phe. After chasing for the indicated times, the medium was removed, and cell proteins were precipitated with TCA. Unincorporated counts at different time points were plotted relative to the initial incorporated radioactivity. **c**, Experimental protocol for **d** and **e** (always in normal medium). mTOR inhibitors (100 nM Torin 1, 300 nM rapamycin) were added at time 0 of chase. **d**, mTOR inhibition with Torin 1 or rapamycin causes rapid and sustained increases in overall protein degradation. Levels of TCA-insoluble pre-labelled proteins remaining at different times were plotted against the initial total incorporated radioactivity. **e**, Treatment with Torin 1 or rapamycin increased both proteasomal and lysosomal proteolysis for at least 19–28 h. Alternatively, the average rates of total proteolysis during the indicated periods were measured by the release of ³H-Phe into medium. The inclusion of concanamycin A (100 nM) to block lysosomal proteolysis allowed specific measurement of proteasomal (concanamycin-A-resistant) and lysosomal (concanamycin-A-sensitive) degradation. Error bars represent s.e.m. ($n = 4$ biological replicates). * $P < 0.05$ (by Student's *t*-test).

demonstrate a prolonged enhancement of proteasome-mediated degradation, which in nutrient-deprived cells must complement the activation of autophagy in providing an endogenous source of amino acids. The major differences between our pulse-chase approach and that of Zhang *et al.* is that we always compare the degradation of the exact same pool of labelled proteins and conducted our assays in complete medium, while they used prolonged serum deprivation of *Tsc2*-null cells. This approach is problematic for pulse-chase studies of the effects of mTORC1 on proteolysis since serum deprivation itself stimulates protein breakdown and mTORC1 alters patterns of synthesis. These differences in methodology may also help explain why Zhang *et al.* did not observe increased proteolysis upon mTORC1 inhibition³ (as shown in Fig. 1 and elsewhere⁴).

Jinghui Zhao¹, Gonzalo A. Garcia² & Alfred L. Goldberg¹

¹Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA.

email: alfred_goldberg@hms.harvard.edu

²Department of Chemistry, Cambridge University, Cambridge CB2 1EW, UK.

Received 26 November 2014; accepted 19 October 2015.

1. Laplante, M. & Sabatini, D. M. mTOR signaling in growth control and disease. *Cell* **149**, 274–293 (2012).
2. He, C. & Klionsky, D. J. Regulation mechanisms and signaling pathways of autophagy. *Annu. Rev. Genet.* **43**, 67–93 (2009).

3. Zhang, Y. *et al.* Coordinated regulation of protein synthesis and degradation by mTORC1. *Nature* **513**, 440–443 (2014).
4. Zhao, J., Zhai, B., Gygi, S. P. & Goldberg, A. L. mTOR inhibition activates overall protein degradation by the ubiquitin proteasome system as well as by autophagy. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1521919112> (2015).
5. Goldberg, A. L. & St John, A. C. Intracellular protein degradation in mammalian and bacterial cells: part 2. *Annu. Rev. Biochem.* **45**, 747–804 (1976).
6. Rock, K. L., Farfan-Arribas, D. J., Colbert, J. D. & Goldberg, A. L. Re-examining class-I presentation and the DRiP hypothesis. *Trends Immunol.* **35**, 144–152 (2014).
7. Zhao, J. *et al.* FoxO3 coordinately activates protein degradation by the autophagic/lysosomal and proteasomal pathways in atrophying muscle cells. *Cell Metab.* **6**, 472–483 (2007).
8. Brault, J. J., Jørgensen, J. G. & Goldberg, A. L. Peroxisome proliferator-activated receptor γ coactivator 1 α or 1 β overexpression inhibits muscle protein degradation, induction of ubiquitin ligases, and disuse atrophy. *J. Biol. Chem.* **285**, 19460–19471 (2010).
9. Piccirillo, R. & Goldberg, A. L. The p97/VCP ATPase is critical in muscle atrophy and the accelerated degradation of muscle proteins. *EMBO J.* **31**, 3334–3350 (2012).
10. Sha, Z. & Goldberg, A. L. Proteasome-mediated processing of Nrf1 is essential for coordinate induction of all proteasome subunits and p97. *Curr. Biol.* **24**, 1573–1583 (2014).

Author Contributions J.Z. and A.L.G. designed experiments, analysed data and prepared this manuscript; J.Z. and G.A.G. performed experiments.

Competing Financial Interests: Declared none.

doi:10.1038/nature16472

Zhang & Manning reply

REPLYING TO: J. Zhao, G. A. Garcia & A. L. Goldberg *Nature* **529**, <http://dx.doi.org/10.1038/nature16472> (2016)

In our study¹, we found that, as a balance to its acute effects on anabolic processes, prolonged activation of mTORC1 increases cellular proteasome content and degradation capacity through the NRF1/NFE2L1 transcription factor. In the accompanying Comment², Zhao *et al.* raise concerns with our methods but do not recognize the importance of distinguishing between mTORC1 and mTORC2 in studying an mTOR-regulated process.

One valid technical point that Zhao *et al.* focus on involves continued label incorporation during early stages of the chase period in a standard pulse-chase assay. Sustained incorporation of ³⁵S-Met is evident in our published data¹, with net loss of protein only being measured following the first chase time point, which was taken relatively late at 8 h. Most importantly, after this initial phase, cells with active mTORC1 were found to degrade protein at higher rates when compared directly to cells with inactive mTORC1, an effect that was blocked by inhibitors of the proteasome but not the lysosome¹. Consistent with the well-established induction of autophagy upon mTORC1 inhibition, the ability of long-term rapamycin treatment to slow protein turnover was even more pronounced in autophagy-deficient cells¹. Given that mTORC1 promotes protein synthesis and, thus, ³⁵S-Met reincorporation, if anything, we might have underestimated the positive effects of mTORC1 on protein degradation in these assays. This technical issue has no bearing on our interpretation of these experiments or their contribution to the central conclusions of our study.

To address whether ³⁵S-Met reincorporation influenced our conclusions, as proposed, we repeated these experiments using ³H-Phe in both a standard 30-min pulse-chase assay and the 20-h labelling experiment preferred by Zhao *et al.* In both cases, the data are similar to that

presented in our paper, but without detectable label reincorporation (Fig. 1). In these experiments, we treated cells with rapamycin only during the chase period to ensure that we are measuring turnover of the same pool of proteins. As published¹, TSC2-deficient cells with activated mTORC1 displayed enhanced rates of protein turnover relative to rapamycin-treated or TSC2-expressing cells, where mTORC1 is inactive. The differences are most evident at the later time points (>5 h), consistent with the NRF1-dependent transcriptional mechanism delineated in our study¹ and discussed in detail elsewhere³. Thus, the discrepancy with the communicated findings of Zhao *et al.* is not a result of the assay, but rather is due to differences in the chosen culture conditions.

A challenging but essential aspect of studying mTOR signalling is the need to clearly distinguish between mTORC1 and mTORC2, two functionally distinct mTOR complexes, in order to gain mechanistic insights into a cellular process⁴. *Tsc1/2*-null cells, which exhibit growth factor-independent activation of mTORC1, have emerged as a powerful genetic tool, but must be cultured and compared to control cells in a manner that properly isolates mTORC1 from mTORC2. Zhao *et al.* present experiments using *Tsc2*-null mouse embryonic fibroblasts, without reconstituted control cells, grown in full serum and treated with Torin 1 or a dose of rapamycin (300 nM) at least 100-fold higher than the IC₅₀ for mTORC1. Under these conditions, one cannot distinguish between effects on mTORC1 or mTORC2, as serum activates both complexes and these doses of inhibitors block both complexes. By contrast, the low-serum growth conditions and treatments we use with these cells are specifically designed to separate effects of mTORC1 from mTORC2, where a clear mTORC1-dependent enhancement of long-term protein degradation is observed (ref. 1 and Fig. 1).

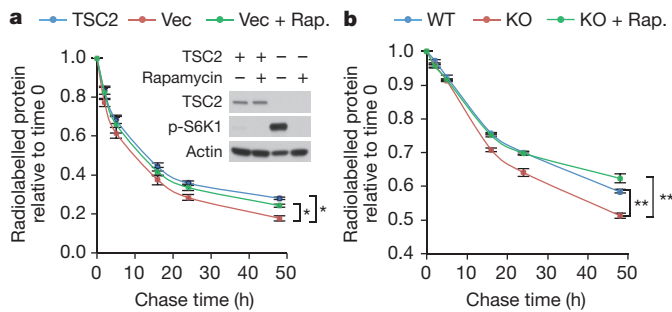


Figure 1 | Active mTORC1 signalling enhances protein degradation in both pulse-chase and long-term labelling assays with ^3H -Phe.
a, *Tsc2*^{-/-} MEFs reconstituted with TSC2 or empty vector (Vec) were cultured for 16 h in medium containing 0.5% serum, with triplicate samples per condition and time point. Cells were pulse-labelled for 30 min with ^3H -Phe, washed three times with cold chase medium containing 2 mM Phe, and chased in medium containing vehicle or rapamycin (20 nM). The amount of radioactivity in TCA-precipitable cellular protein was measured by scintillation, with the rate of protein degradation shown as the fraction of radiolabelled protein, relative to time 0, remaining over time. Immunoblot inset: mTORC1 signalling under these experimental conditions, as detected with phospho (p)-S6K1 (T389), at the 16-h time point. **b**, Cells were treated as in **a**, except radiolabelled for 20 h with ^3H -Phe and washed for 2 h with cold chase medium containing 2 mM Phe before measuring degradation of the remaining protein over time. Except for the serum and rapamycin concentrations, this assay is identical to that described in the accompanying Comment². KO, knockout; WT, wild type. Data are a composite of two independent experiments with three biological replicates each and are expressed as mean \pm s.d. * $P < 0.05$, ** $P < 0.01$.

We stand by our conclusions that mTORC1 activation promotes the production of proteasomes leading to enhanced proteasome-mediated protein turnover through an increase in NRF1 (refs 1, 3), a

transcription factor now established in multiple independent studies to control cellular proteasome levels^{1,3-9}.

The authors provide this response on behalf of all of the authors of our published manuscript¹, and each author has reviewed and approved of its contents.

Yinan Zhang¹ & Brendan D. Manning¹

¹Department of Genetics and Complex Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, USA.
 email: bmannings@hsph.harvard.edu

1. Zhang, Y. *et al.* Coordinated regulation of protein synthesis and degradation by mTORC1. *Nature* **513**, 440–443 (2014).
2. Zhao, J., Garcia, G. A. & Goldberg, A. L. Control of proteasomal proteolysis by mTOR. *Nature* **529**, <http://dx.doi.org/doi:10.1038/nature13492> (2015).
3. Zhang, Y. & Manning, B. D. mTORC1 signaling activates NRF1 to increase cellular proteasome levels. *Cell Cycle* **14**, 2011–2017 (2015).
4. Shimobayashi, M. & Hall, M. N. Making new contacts: the mTOR network in metabolism and signaling crosstalk. *Nature Rev. Mol. Cell Biol.* **15**, 155–162 (2014).
5. Radhakrishnan, S. K. *et al.* Transcription factor Nrf1 mediates the proteasome recovery pathway after proteasome inhibition in mammalian cells. *Mol. Cell* **38**, 17–28 (2010).
6. Steffen, J., Seeger, M., Koch, A. & Kruger, E. Proteasomal degradation is transcriptionally controlled by TCF11 via an ERAD-dependent feedback loop. *Mol. Cell* **40**, 147–158 (2010).
7. Radhakrishnan, S. K., den Besten, W. & Deshaies, R. J. p97-dependent retrotranslocation and proteolytic processing govern formation of active Nrf1 upon proteasome inhibition. *Elife* **3**, e01856 (2014).
8. Lee, C. S., Ho, D. V. & Chan, J. Y. Nuclear factor-erythroid 2-related factor 1 regulates expression of proteasome genes in hepatocytes and protects against endoplasmic reticulum stress and steatosis in mice. *FEBS J.* **280**, 3609–3620 (2013).
9. Sha, Z. & Goldberg, A. L. Proteasome-mediated processing of Nrf1 is essential for coordinate induction of all proteasome subunits and p97. *Curr. Biol.* **24**, 1573–1583 (2014).

doi:10.1038/nature16473

FORUM Evolution

A sisterly dispute

Which phylum first branched off from the animal phylogenetic tree is a contested issue. A new analysis challenges the proposal that comb jellies are the sister group to all other animals, and emphasizes a 'sponges-first' view. Three evolutionary biologists weigh up the evidence.

THE PAPER IN BRIEF

- There are five major branches of animals: Porifera (sponges), Cnidaria (jellyfishes, corals and related species), Ctenophora (comb jellies), Placozoa (*Trichoplax*) and Bilateria (all other animal phyla).
- Writing in *Proceedings of the National Academy of Sciences*, Pisani *et al.*¹ reanalyse some existing data and support the case

that Porifera are the sister group to all other animals.

- The authors propose that this conclusion fits with the fact that sponges lack features present in the other phyla, such as a nervous system and muscles.
- However, other recent genomic analyses have suggested that the more complex Ctenophora are the sister group.

Fighting over a comb

MAXIMILIAN J. TELFORD

The Ctenophora, also known as comb jellies or sea gooseberries, are a phylum of beautiful marine animals. Several molecular phylogenetic studies^{2–4} have made the extraordinary claim that Ctenophora, despite sharing 'advanced' characteristics — such as muscles, nerves and epithelial tissues — with Bilateria and Cnidaria, are more distantly related to these groups than are the simple Porifera, which lack these features.

In contrast to the conventional 'Porifera-sister' hypothesis, the Ctenophora-sister tree, in which the Ctenophora are the sister group of all other animals, implies either that muscles, nerves and epithelia evolved independently in two animal lineages (in Ctenophora and in the ancestor of Bilateria and Cnidaria), or that these characteristics are ancient and have been lost by sponges (Fig. 1). Considering these surprising implications, it was inevitable that the Ctenophora-sister idea would be highly controversial^{5–7}.

Critics of the original studies suggest that they were affected by errors in tree reconstruction caused by the use of inadequate phylogenetic models. Pisani *et al.*¹ investigate this possibility by reanalysing several published data sets comprising concatenated alignments of many genes^{2–5}, and a data matrix recording the presence or absence of more than 23,000 genes in different animal species³. The authors

present three lines of evidence to suggest that the Ctenophora-sister trees are an artefact.

First, recognizing that tree-reconstruction errors can stem from a poor fit between an evolutionary model and real data, they use the statistical technique of cross-validation to show that a 'site-heterogeneous' model⁸ best fits the data. These CAT models (named for modelling multiple categories of site) avoid the assumption of a homogeneous process of amino-acid substitution across sites within genes, and they have repeatedly been shown to outperform site-homogeneous models. The CAT model provided the best fit to the published data and supported the Porifera-sister hypothesis; less-well-fitting models better suit a Ctenophora-sister situation.

The authors' second approach considers the common artefact of long-branch attraction (LBA), which could cause the long ctenophore branch to be attracted downwards towards the long branch leading to the non-animal species that form the root of the tree⁹ — these groups are referred to as outgroups. The authors reason that, the longer the branch leading to the outgroups, the higher the likelihood of LBA. Removing the most distant outgroups shifts support from the Ctenophora-sister to the Porifera-sister picture, suggesting that the Ctenophora-sister result is an effect of LBA.

Third, the authors show that previous analyses of the gene presence-or-absence matrix, cited as an independent source of evidence supporting the Ctenophora-sister tree³, were affected by an unintentional bias: genes present in only one species were not recorded in the original matrix, and the prior existence of genes that have been lost in all sampled

species was ignored. These exclusions lead to a systematic underestimation of the likelihood of a gene being lost in any given species. The effect is that missing genes (such as those absent from ctenophores) are more likely to be interpreted as never having existed than as having been lost during evolution — effectively pushing species that have lost many genes towards the root of the tree. When this known problem with presence-absence data¹⁰ is corrected, the result is a much more credible tree (the previous work had several odd features), and independent support for the Porifera-sister hypothesis.

The position of Ctenophora has proved so difficult to determine because it involves three confounding circumstances: the main animal groups separated more than 540 million years ago during the Precambrian period, meaning that the phylogenetic signal supporting any clade is faint; Ctenophora was a fast-evolving taxon that seems to have lost features along the way; and there is a lack of informative intermediates on a long branch leading to the recent radiation of modern ctenophore species. Careful use of models designed to avoid the worst effects of these problems, and experiments designed to detect these issues, are essential for us to understand the true affinities of this phylum. The analyses that fit these requirements support the Porifera-sister camp, providing a parsimonious interpretation of morphological evolution.

Maximilian J. Telford is in the Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK.
e-mail: m.telford@ucl.ac.uk

Methodological misconceptions

LEONID L. MOROZ & KENNETH M. HALANYCH

Each new analysis of early animal phylogeny fuels debate about animal origins and the parallel evolution of animal complexity. Pisani *et al.*¹ have used phylogenetic methods that we

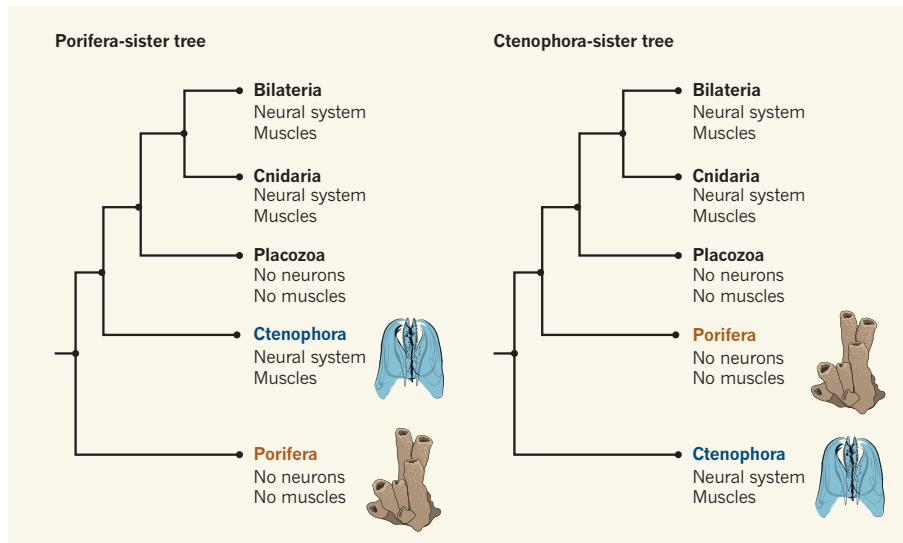


Figure 1 | The animal tree. There is debate over whether Porifera (sponges) or Ctenophora (comb jellies) were the first group to branch off from the animal tree. **a**, Porifera have typically been considered the simplest group, because they lack muscles and neurons, which are present in Ctenophora, Cnidaria (jellyfishes, corals and related species) and Bilateria (other animals). **b**, However, some phylogenetic analyses have suggested that Porifera are more closely related than Ctenophora to Bilateria and Cnidaria. (Placozoa, generally classified as the single species *Trichoplax adhaerens*, are consistently placed as the sister group of Cnidaria and Bilateria, but also lack neurons and muscles.)

consider unproven to reanalyse a small fraction of published data sets^{2–4}, and they reject the hypothesis that ctenophores are the sister group to other animals. But the authors' assertions are based on incomplete analyses and inappropriate assumptions, and their conclusion is at odds with other studies, including two recent analyses^{11,12}.

A key point is that convergent evolution of complex characteristics is compatible with both the Ctenophora-sister and Porifera-sister hypotheses^{4,13}. Morphological, functional or molecular complexity does not correlate with phylogeny. More than 550 million years of animal evolution has produced countless examples of independent gains and losses of complex structures¹⁴, and hypotheses based on the idea that morphologically simpler animals came first should not be blindly accepted, nor should complex features be treated as single characteristics in phylogenetic analyses.

Convergent evolution can be surmised if the system in question differs in structure, molecular composition and mechanisms. This case is met for genes expressed in neurons and in smooth and striated muscles in ctenophores versus those in cnidarians and bilaterians^{15,16}, implying that these features were not shared in the common ancestor. Integrative multiple-trait analysis and direct microanalytical measurements have demonstrated that ctenophores possess distinct neural machinery⁴, and there are no known pan-neuronal or neural genes that are shared by all animals¹⁷.

In fact, neural systems should not be considered as a single character in evolutionary reconstructions. They are composed of highly distinct cell populations with

different histories and origins, and relationships between these different cell lineages across phyla must be reconstructed to decipher neuronal genealogies. Furthermore, the placement of the nerveless Placozoa in Pisani and colleagues' tree challenges the assumption of a single origin of the nervous system (Fig. 1). Placing either sponges or ctenophores as sister to all other animals does not alter the possible scenario of independent origins of neural systems^{13,15} nor of muscles in ctenophores^{4,16}.

Several uncertainties about phylogenomic analyses are also illustrated by Pisani and colleagues' approaches. The authors assert that CAT models, which correct for variation between amino-acid positions⁸, outperform models used in other studies^{2–4}. However, other studies^{2,4} applied a different model to each partition in the data set, whereas Pisani and colleagues' validation approach uses only a single model across the entire data set. The authors' justification for superiority of CAT models is also based on assumptions about the 'true tree', including the placement of sponges, leading to a circular argument⁷. Moreover, CAT models are computationally demanding, and can produce conflicting results^{1,7,11,12} or fail to run to completion^{1–4}, and Pisani *et al.* discuss several incomplete analyses as providing support for the Porifera-sister hypothesis.

The authors chose to use only the taxon most closely related to animals (unicellular and colonial choanoflagellates) as the outgroup on their animal tree. It is not clear whether this assumption was made because it yielded a Porifera-sister tree or in response to objective criteria. The approach used does not allow for

the control of potential artefacts, such as the fact that choanoflagellates and some sponges have markedly different amino-acid composition from most animals, which could lead to incorrect rooting of the animal tree. Prior analyses^{2,4} that have thoroughly explored out-group choice with objective criteria support the ctenophore-sister hypothesis.

Finally, it is worth considering other lines of evidence alongside phylogenetic reconstructions. Morphological characteristics that seem to unite the choanocyte cells of sponges with choanoflagellates may not actually be homologous¹⁸. And palaeontological evidence for the appearance of sponges during the Cryogenian period (around 850 million to 635 million years ago) has been challenged¹⁹. Confirmed sponge fossils appeared only with most other animal groups, near the transition from the Ediacaran to the Cambrian period, around 540 million years ago¹⁹. The Porifera-sister hypothesis, as promoted by Pisani *et al.*, needs further scrutiny from both phylogenomic and other standpoints. By contrast, analyses based on objective criteria favour the Ctenophore-sister hypothesis and parallel evolution of animal complexities across phyla. ■

Leonid L. Moroz is at the Whitney Laboratory for Marine Bioscience, University of Florida, St Augustine, Florida 32080, USA, and in the Department of Neuroscience and at the McKnight Brain Institute, University of Florida, Gainesville. **Kenneth M. Halanych** is in the Department of Biological Sciences, Auburn University, Auburn, Alabama 36849, USA.
e-mails: moroz@whitney.ufl.edu; ken@auburn.edu

- Pisani, D. *et al.* *Proc. Natl Acad. Sci. USA* **112**, 15402–15407 (2015).
- Whelan, N. V., Kocot, K. M., Moroz, L. L. & Halanych, K. M. *Proc. Natl Acad. Sci. USA* **112**, 5773–5778 (2015).
- Ryan, J. F. *et al.* *Science* **342**, 1242592 (2013).
- Moroz, L. L. *et al.* *Nature* **510**, 109–114 (2014).
- Jékely, G., Paps, J. & Nielsen, C. *EvoDevo* **6**, 1 (2015).
- Marlow, H. & Arendt, D. *Curr. Biol.* **24**, R757–R761 (2014).
- Philippe, H. *et al.* *Curr. Biol.* **19**, 706–712 (2009).
- Lartillot, N., Lepage, T. & Blanquart, S. *Bioinformatics* **25**, 2286–2288 (2009).
- Felsenstein, J. *Syst. Biol.* **27**, 401–410 (1978).
- Felsenstein, J. *Evolution* **46**, 159–173 (1992).
- Borowiec, M. L., Lee, E. K., Chiu, J. C. & Plachetzki, D. C. *BMC Genomics* **16**, 987 (2015).
- Chang, E. S. *et al.* *Proc. Natl Acad. Sci. USA* **112**, 14912–14917 (2015).
- Moroz, L. L. *Commun. Integr. Biol.* **7**, e993269 (2014).
- Moore, J. & Willmer, P. *Biol. Rev. Camb. Phil. Soc.* **72**, 1–60 (1997).
- Moroz, L. L. & Kohn, A. B. *Phil. Trans. R. Soc. B* **371**, 20150041 (2016).
- Steinmetz, P. R. H. *et al.* *Nature* **487**, 231–234 (2012).
- Moroz, L. L. & Kohn, A. B. *Integr. Comp. Biol.* **55**, 1028–1049 (2015).
- Mah, J. L., Christensen-Dalsgaard, K. K. & Leys, S. P. *Evol. Dev.* **16**, 25–37 (2014).
- Antcliffe, J. B., Callow, R. H. T. & Brasier, M. D. *Biol. Rev.* **89**, 972–1004 (2014).

Another energy source for the geodynamo

Magnesium is not usually considered to be a constituent of Earth's core, but its presence there has now been proposed to explain an ongoing enigma — the identity of the energy sources that drive our planet's magnetic field. [SEE LETTER P.387](#)

BRUCE BUFFETT

Turbulent flow in Earth's liquid-iron core generates the planet's magnetic field through a process known as the geodynamo. This process is sustained by energy drawn from the core as it slowly cools¹. Thermal convection is thought to be crucial, but revised estimates^{2,3} of thermal conductivity in liquid iron at high pressure have called into question the adequacy of the commonly cited energy sources¹. On page 387 of this issue, O'Rourke and Stevenson⁴ propose a solution to this energy crisis. They argue that if magnesium had dissolved in the liquid iron at high temperature when the core formed, then subsequent precipitation of magnesium-bearing minerals on cooling would be an important source of energy. The authors' theory warrants a serious reassessment of magnetic-field generation in other rocky (terrestrial) planets.

Sustaining a magnetic field is difficult for a terrestrial planet. Creeping flow of the planet's rocky shell (mantle) restricts heat loss from the underlying core. By comparison, the liquid-iron core is an efficient thermal conductor. Thermal convection in the core ceases when heat flow into the mantle falls below the core's capacity to deliver this heat by conduction

alone, so high thermal conductivity may push the threshold for thermal convection beyond reach. In this scenario, turbulent flow in the core is driven mainly by buoyancy effects due to variations in the abundance of core constituents — as the core cools, some of the iron solidifies and accumulates on the solid inner core, leaving lighter elements in the liquid outer core and thus causing convection¹.

A problem with this conventional view of the geodynamo's energy sources emerges when we extrapolate back in time. Before the inner core formed (possibly less than 1 billion years ago¹), the only energy source was thermal convection. But current estimates of iron's thermal conductivity suggest that the heat flow required to sustain such convection at that time was extremely high. Even if this heat flow was feasible, an implausibly high core temperature would be needed to sustain it over geological time. Despite these difficulties, Earth has somehow maintained a magnetic field for at least the past 3.4 billion years⁵.

O'Rourke and Stevenson address this quandary by proposing a new energy source. They suggest that magnesium can enter the core to form an iron alloy, even though it is normally considered to be nearly insoluble in liquid iron. Other alloying elements are more commonly proposed⁶ to explain why estimates

of the core's density, based on seismic data, are less than that of pure iron. But theoretical predictions⁷ and some experiments (see ref. 8, for example) suggest that magnesium can dissolve in liquid iron at sufficiently high temperatures. The authors argue that, because of its insolubility in iron, magnesium would probably become supersaturated as the core cools. The subsequent precipitation of magnesium-bearing minerals would leave behind a residual liquid enriched in iron, providing a compositional buoyancy that would drive fluid flow (Fig. 1).

Two factors determine whether magnesium precipitation is a substantial energy source. First, the amount of magnesium that dissolved in liquid iron during core formation must be sufficient to meet the energy demands of the geodynamo. Second, the temperature dependence of magnesium's solubility must be strong enough to promote supersaturation of magnesium with only a modest temperature decrease of the core (possibly just several hundred kelvin). Otherwise, a delay in the onset of magnesium precipitation could shut off the energy source in the past or present.

O'Rourke and Stevenson tackled the first issue using previously reported experimental data⁹ that describe the partitioning of elements between liquid iron and silicate melts, a mixture that represents the composition and state of the mantle during core formation. These data allowed them to estimate the concentration of magnesium, oxygen and silicon in liquid iron, as well as the abundance of siderophile elements (those that have an affinity for iron: nickel, cobalt, chromium, vanadium, niobium and tantalum) in the silicate melt, for two models of core formation. They then used a computational technique (a Monte Carlo method) to assess the average temperature and pressure conditions of core formation in the two models, by sampling many possible outcomes.

Crucially, the researchers could account for the observed abundances of siderophiles in the mantle⁹ by using a model that permits a small fraction of the core to equilibrate with silicate melt at high temperatures (greater than 5,000 K). Many of the Monte Carlo outcomes for this model are also compatible with seismological constraints on the abundance of light elements in the core⁶. These results include outcomes in which the liquid core contains 1–2% magnesium by weight. In other words, enough magnesium to power the geodynamo could have been dissolved in the early liquid core without violating known constraints on the composition of the core and mantle^{6,9}.

However, uncertainties prevent a definitive assessment of magnesium precipitation. Some of the realizations sampled by the Monte Carlo method delay magnesium precipitation into the distant future, whereas others permit precipitation much earlier; early and

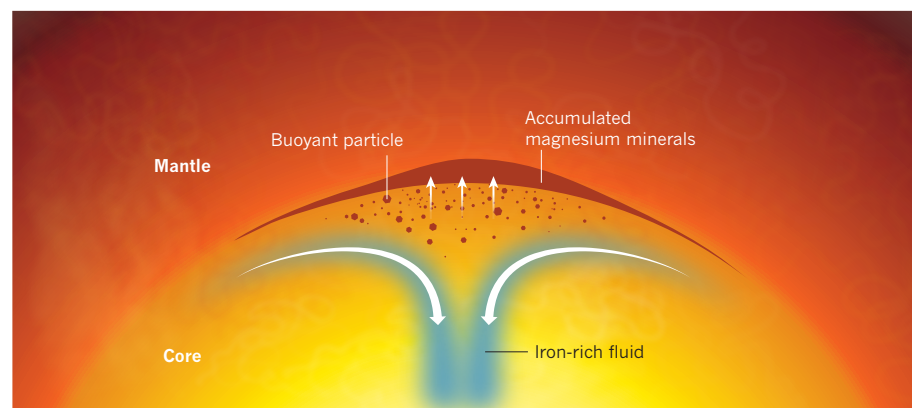


Figure 1 | Possible processes at the boundary between Earth's liquid core and rocky mantle. O'Rourke and Stevenson⁴ propose that magnesium dissolved in liquid iron at high temperature when the core formed. Subsequent precipitation of magnesium-containing minerals would produce buoyant solid particles that would rise through the liquid iron and accumulate at the top of the core. The residual liquid would be denser than the surrounding fluid because it is enriched in iron, and would therefore sink. The resulting vigorous convection would stir the core and generate Earth's magnetic field, at modest rates of core cooling.

abundant precipitation is required to provide an effective solution to the geodynamo energy crisis. Much of the uncertainty derives from the experimental estimates of element partitioning between iron and silicate, particularly at high temperature. For example, the present work used information from a single set of experiments to derive magnesium's solubility⁸. More work is clearly required to address these uncertainties, but the potential contribution of magnesium precipitation to the geodynamo provides plenty of motivation to improve our current knowledge.

Magnesium precipitation would produce a buoyant solid that rises to the top of the core¹⁰. The dense, iron-rich residual fluid would also contribute to vigorous convection, offering ample energy for the geodynamo at relatively

modest cooling rates. Such low cooling rates would allow warm fluid to accumulate at the top of the core, although convection due to magnesium precipitation might mix this warm fluid back to the core's interior. Further complications are suggested by experimental evidence⁹ that the core's liquid is not saturated with oxygen and silicon, indicating that these elements might transfer into the core from the mantle. The potential for two-way transfer across the core–mantle boundary in the light of O'Rourke and Stevenson's theory is likely to send Earth scientists back to the drawing board. ■

Bruce Buffett is in the Department of Earth and Planetary Science, University of California, Berkeley, Berkeley,

California 94720-4767, USA.
e-mail: bbuffett@berkeley.edu

1. Nimmo, F. in *Treatise on Geophysics* 2nd edn (ed. Schubert, G.) 31–55 (Elsevier, 2015).
2. Pozzo, M., Davies, C., Gubbins, D. & Alfè, D. *Nature* **485**, 355–358 (2012).
3. Gomi, H. & Hirose, K. *Phys. Earth Planet. Inter.* **247**, 2–10 (2015).
4. O'Rourke, J. G. & Stevenson, D. J. *Nature* **529**, 387–389 (2016).
5. Tarduno, J. A. *et al. Science* **327**, 1238–1240 (2010).
6. Badro, J., Côté, A. S. & Brodholt, J. P. *Proc. Natl Acad. Sci. USA* **111**, 7542–7545 (2014).
7. Wahl, S. M. & Militzer, B. *Earth Planet. Sci. Lett.* **410**, 25–33 (2015).
8. Takafuji, N., Hirose, K., Mitome, M. & Bando, Y. *Geophys. Res. Lett.* **32**, L06313 (2005).
9. Fischer, R. A. *et al. Geochim. Cosmochim. Acta* **167**, 177–194 (2015).
10. Buffett, B. A., Garnero, E. J. & Jeanloz, R. *Science* **290**, 1338–1342 (2000).

CANCER

Bet on drug resistance

Inhibitors of the BET bromodomain proteins are promising cancer therapeutics, but tumour cells are likely to become resistant to these drugs. Anticipated mechanisms of resistance have now been described. [SEE LETTER P.413](#)

JEFF SETTLEMAN

Many modern cancer drugs target mutationally activated proteins, but this treatment strategy has limitations. Only a relatively small number of mutations are seen recurrently across human tumours¹, and drug resistance develops rapidly². Targeting the epigenome³ — the chemically modified form of DNA, and of associated histones and other proteins that facilitate the packaging of DNA as chromatin, all of which influence gene expression — is one of the alternative approaches being explored. Along with two papers^{4,5} published in *Nature* last year, a paper⁶ on page 413 of this issue provides some insight into the potential of epigenome-targeting drugs called BET inhibitors, and outlines the mechanisms by which tumours might become resistant to these drugs.

It has long been recognized⁷ that tumour cells have distinct epigenomic features, which can lead to the overproduction of cancer-promoting transcription factors such as MYC. Transcription factors are challenging therapeutic targets, because they lack structures that can be readily targeted with drugs. But developments in our understanding of the epigenome-regulating factors that influence gene expression, many of which seem to be 'druggable', have provided a potential way to sidestep this hurdle.

Among these factors is the bromodomain protein family, which includes the BET

subfamily⁸ (BRD2, 3, 4 and T). BET proteins contain two bromodomains, each with small pockets. These pockets bind to histones that have been tagged with acetyl groups, enabling BET proteins to recruit the cell's transcriptional machinery to specific sites in the genome to regulate gene expression. BET subfamily members such as BRD4, which can regulate MYC gene transcription, have been implicated in various tumours (particularly in cancers of the blood) and are therefore candidate targets for therapy⁸.

A few years ago, the first of several small-molecule BET inhibitors (JQ1) was discovered, and shown⁸ to effectively disrupt cancer-cell proliferation. This effect seemed to reflect inhibition of BET-mediated regulation of MYC expression. Early clinical trials of BET inhibitors in leukaemia and lymphoma have been encouraging. Investigators are now seeking other disease contexts in which these inhibitors might work, and predicting the acquired resistance mechanisms that will inevitably arise.

The two 2015 studies^{4,5} converge on a potential mechanism of resistance to BET inhibition in acute myelogenous leukaemia (AML). In the first, Rathert *et al.*⁴ screened mouse AML cells for chromatin-modifying factors that are required for AML-cell survival. They confirmed that AML cells need Brd4, and identified several other factors for which inhibition confers AML-cell resistance to JQ1. In AML cells that were JQ1-resistant, the authors observed changes in specific epigenome features in DNA enhancer regions,

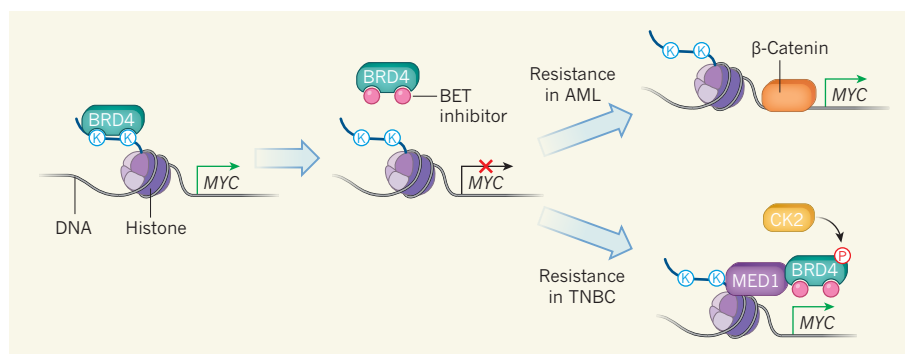


Figure 1 | Circumventing BET inhibition. The BET protein BRD4 can bind to acetyl groups (K) on histone proteins around which DNA is packaged as chromatin. BRD4 recruits the cell's transcriptional machinery, upregulating expression of the cancer-promoting MYC gene. Treatment with BET inhibitors can prevent BRD4–chromatin binding, stalling MYC transcription, but cancer cells rapidly develop drug resistance. Rathert *et al.*⁴ and Fong *et al.*⁵ report that, in acute myelogenous leukaemia (AML), drug resistance is conferred by activation of the Wnt-signalling pathway, which leads to DNA binding and MYC activation by the protein β -catenin. By contrast, Shu *et al.*⁶ find that resistance in triple-negative breast cancer (TNBC) arises owing to activation of the casein kinase 2 (CK2) enzyme. CK2 phosphorylates (P) BRD4, allowing BRD4 to bind to the transcriptional regulator protein MED1 to activate MYC.

which regulate gene expression. These changes meant that *MYC* gene expression could be activated without Brd4.

Rathert *et al.* also found that genes associated with the Wnt-signalling pathway, a known driver of tumour development, were upregulated in resistant cells. Wnt activation was sufficient to promote JQ1 resistance, possibly by driving the transcription of *MYC* at an enhancer generated specifically in the resistant cells (Fig. 1). Finally, the authors found the same mechanism of resistance to JQ1 in some other cancer types and in blood cells taken from people with leukaemia. Together, their data suggest that the usefulness of BET inhibitors could be expanded by combining them with Wnt-pathway inhibitors.

Taking a different approach, Fong *et al.*⁵ rendered mouse AML cells resistant to BET inhibition by continuously exposing these cells to another BET inhibitor, eventually yielding drug-resistant clonal populations. This experiment also showed that Wnt activation has a role in drug resistance. Furthermore, the resistant cells had features of stem cells, suggesting that the AML cancer-stem-cell population, or a subset thereof, does not respond to BET inhibitors.

The Wnt pathway has previously been shown to be involved in drug resistance in AML cancer stem cells⁹. Moreover, the drug-resistant nature of cancer stem cells is well established¹⁰. However, Rathert and colleagues did not find evidence that the resistant AML cells had stem-cell features — a distinction between the two reports.

In the current study, Shu *et al.*⁶ explored BET inhibition in human breast cancer. By profiling a panel of breast-cancer cell lines, they observed that one cancer subtype — 'triple-negative' breast cancer — was sensitive to BET inhibition. Like Fong *et al.*, the authors modelled acquired resistance to BET inhibition by culturing sensitive triple-negative cells in JQ1, and then characterized emergent resistant cells. Resistant cells remained dependent on BRD4, but this dependence did not involve the protein's bromodomains.

A widely active transcriptional regulator protein called MED1 bound more tightly to BRD4 in resistant cells than in sensitive cells. The authors attributed this tighter binding to increased BRD4 phosphorylation mediated by the enzyme casein kinase 2 (CK2). The binding gave rise to bromodomain-independent, BRD4-mediated transcriptional activation of *MYC*, among other genes (Fig. 1). These data suggest that using a combination of CK2 and BET inhibitors to treat triple-negative breast cancer might prevent drug resistance.

Although many previous studies have demonstrated the efficacy of drugs against triple-negative breast cancer in animal and cell-based models, it is worth noting that these drugs have so far failed to combat tumours in people. As such, optimism should be tempered.

Collectively, these three reports show that

BET inhibitors might have a broader potential than had previously been realized. They also highlight the possibility that BET inhibitors could be used in combination with other drugs to overcome both innate and acquired drug resistance. Although the reported resistance mechanisms seem to reflect an adaptation to drug pressure, the root cause of resistance remains unknown. Does a specific mutation cause Wnt or CK2 activation, or are these adaptive changes that drive resistance through reversible epigenetic mechanisms? A complete mechanistic understanding of resistance remains to be defined.

It is important to note that clinical inhibitors of Wnt or CK2 have yet to be developed. Therefore, the hypotheses that emerge from these studies cannot be tested in the clinic. Nonetheless, these three reports provide a good foundation on which to build a better

understanding of mechanisms of resistance that should be anticipated in the clinic. ■

Jeff Settleman is at Calico Life Sciences, South San Francisco, California 94080, USA.
e-mail: settleman@calicolabs.com

1. Klijn, C. *et al.* *Nature Biotechnol.* **33**, 306–312 (2015).
2. Lackner, M. R., Wilson, T. R. & Settleman, J. *Future Oncol.* **8**, 999–1014 (2012).
3. Baylin, S. B. & Jones, P. A. *Nature Rev. Cancer* **11**, 726–734 (2011).
4. Rathert, P. *et al.* *Nature* **525**, 543–547 (2015).
5. Fong, C. Y. *et al.* *Nature* **525**, 538–542 (2015).
6. Shu, S. *et al.* *Nature* **529**, 413–417 (2016).
7. Eilers, M. & Eisenman, R. N. *Genes Dev.* **22**, 2755–2766 (2008).
8. Qi, J. *Cold Spring Harb. Perspect. Biol.* **6**, a018663 (2014).
9. Wang, Y. *et al.* *Science* **327**, 1650–1653 (2010).
10. Singh, A. & Settleman, J. *Oncogene* **29**, 4741–4751 (2010).

This article was published online on 6 January 2016.

EPIDEMIOLOGY

A global picture of melioidosis

Comprehensive mapping and modelling have estimated global deaths from the bacterial disease melioidosis to be comparable to deaths from measles and substantially greater than those from dengue or leptospirosis.

BART J. CURRIE & MIRJAM KAESTLI

The bacterium *Burkholderia pseudomallei* is found in soil and water and causes the disease melioidosis in humans and animals. It was upgraded to a Tier 1 Select Agent by the US Centers for Disease Control and Prevention in 2012 — the designation given to pathogens considered of highest risk. This concern is based on a mortality rate of up to 40%, the fact that *B. pseudomallei* can be easily aerosolized in laboratory studies, its intrinsic resistance to standard antimicrobial agents and the lack of a vaccine. Writing in *Nature Microbiology*, Limmathurotsakul *et al.*¹ compile global literature on cases of melioidosis in humans and animals, and on the environmental presence of *B. pseudomallei*. The authors combine these data with environmental and demographic parameters in a modelling framework to predict countries in which melioidosis is probably endemic, and to estimate the global burden of the disease.

Melioidosis was first described in 1912 as a septicaemic disease in morphine addicts in Yangon, Myanmar. It is now recognized as a major cause of fatal pneumonia and sepsis in Thailand, Malaysia, Singapore and northern Australia². Limmathurotsakul *et al.* report that melioidosis is, in fact, known to be endemic

in 48 countries spanning the tropics and including nations in East and South Asia, the Middle East, Africa, Latin America, the Caribbean and the Pacific. Furthermore, the authors' modelling suggests that melioidosis is severely under-reported in most of these countries and is probably endemic in another 34 countries that so far have had no confirmed cases.

The authors' estimate for total global human cases for 2015 was 165,000 (a 95% credible interval of 68,000–412,000), with 89,000 deaths (95% credible interval 36,000–227,000). As the authors point out, the global mortality estimates are comparable to those due to measles and much higher than those due to leptospirosis and dengue infection. That melioidosis is so badly underdiagnosed is not surprising, because most cases occur in resource-poor countries that have large rural populations and limited or no capacity for microbiological laboratory diagnosis — most crucially, the ability to culture blood samples and identify recovered bacteria.

Limmathurotsakul and colleagues' comprehensive investigation resulted in a database of 22,338 geographically located records of cases of human and animal melioidosis and the presence of environmental *B. pseudomallei*, reported between 1910 and 2014. The listing of 48 endemic countries is



Figure 1 | Environments conducive to *Burkholderia pseudomallei* success. The bacterial agent of the disease melioidosis has been cultivated from environments as diverse as rice paddies in Cambodia (left) and waterholes in northern and central Australia (right).

based on confirmed cases or the presence of environmental *B. pseudomallei*, but the predictions of global case numbers, deaths and endemicity are based on modelling using these data.

To predict the global population at risk, the authors used a statistical model known as a boosted regression tree to link various environmental conditions to the confirmed presence of *B. pseudomallei*. As the authors note, such models have been used to map the global burden of dengue infection³. The authors then used their model to assess the environmental suitability for *B. pseudomallei* globally; this predicted the bacterium to be ubiquitous throughout the tropics. The global incidence of melioidosis cases was predicted using a negative binomial model that assessed the association between melioidosis incidence rates from 16 reports from endemic locations and the predicted environmental suitability for *B. pseudomallei*, as well as further population-based parameters, including the prevalence of diabetes (the major risk factor for melioidosis).

Although *B. pseudomallei* can survive and thrive in diverse environments (Fig. 1), this does not mean that it is actually present in all the predicted locations. Studies of microbial biogeography have revealed the complexity of global dispersal patterns, with variable influences of habitat, geographical boundary, historical timescale, anthropogenic impact and organism dispersal ability⁴. Each of these factors has relevance for *B. pseudomallei*. Phylogeographic reconstruction of *B. pseudomallei* genomes supports an Australian origin for the species from an ancestral strain of *Burkholderia* in the local environment, with possibly a single introduction event into south-east Asia that was estimated to have occurred during the last glacial period (between 16,000 and 225,000 years ago)⁵. More-recent

dissemination to Africa and the Americas is hypothesized, but the timelines and modes of dispersal are unclear.

The expansion of known endemic locations for melioidosis in recent years may mostly reflect the fact that improved diagnostics have unmasked a long-standing presence of *B. pseudomallei* in the local environment. Alternatively, there may have been substantial dissemination as a result of increasing human, animal, plant and soil movements. Global warming may also hasten expansion of the endemic boundaries of the disease⁶. Reports from Brazil, Madagascar and Papua New Guinea suggest that melioidosis may be restricted to regional hotspots in some endemic countries. This may reflect under-ascertainment or environmental determinants yet to be elucidated, or that the bacterium hasn't yet dispersed widely in those countries.

Another consideration is that different melioidosis-endemic locations may vary in their specific ecological niches for *B. pseudomallei*, with the bacterium potentially providing a biodefence function for local cohabiting plant species. In addition, some introduced plants, such as pasture grasses, have been shown to become heavily colonized by *B. pseudomallei*⁷. Modelling environmental parameters from one region may not necessarily predict findings in another. Limmathurotsakul and colleagues' environmental-suitability model did show maximum rainfall rather than average rainfall or temperature to be the most important model contributor. This reflects the distinctive seasonal rainfall pattern of a tropical wet-dry climate, which is seen in the regions of Thailand and Australia that have the highest documented incidences of melioidosis, but not in Singapore. Salinity has been reported to be a negative predictor of *B. pseudomallei* presence⁸, but this is in contrast to the model's findings.

Other uncertainties surround the disease itself. It is not clear what proportion of melioidosis cases result from inhalation compared with infection through the skin⁶. Epidemiological reports suggest that increased inhalation-derived infections occur during severe weather events such as cyclones and typhoons⁶, and ingestion of *B. pseudomallei* from unchlorinated water seems to have more impact than previously thought⁹.

As well as predicting the endemicity of melioidosis in many countries in which the disease has not yet been recorded, Limmathurotsakul and colleagues' modelling predicts incidence rates of, for example, more than 50,000, 20,000 and 13,000 annual undiagnosed cases in India, Indonesia and Nigeria, respectively. Targeted surveillance, together with support for improved regional microbiology facilities, are needed to reveal the accuracy of these predictions¹⁰. ■

Bart J. Currie and Mirjam Kaestli are in the Global and Tropical Health Division, Menzies School of Health Research, Charles Darwin University, Darwin, Northern Territory 0810, Australia. **B.J.C.** is also in the Infectious Diseases Department, Royal Darwin Hospital. e-mails: bart.currie@menzies.edu.au; mirjam.kaestli@menzies.edu.au

1. Limmathurotsakul, D. et al. *Nature Microbiol.* **1**, 15008 (2016).
2. White, N. J. *Lancet* **361**, 1715–1722 (2003).
3. Bhatt, S. et al. *Nature* **496**, 504–507 (2013).
4. Martiny, J. B. H. et al. *Nature Rev. Microbiol.* **4**, 102–112 (2006).
5. Pearson, T. et al. *BMC Biol.* **7**, 78 (2009).
6. Currie, B. J. *Semin. Respir. Crit. Care Med.* **36**, 111–125 (2015).
7. Kaestli, M. et al. *Environ. Microbiol.* **14**, 2058–2070 (2012).
8. Inglis, T. J. J. & Sagripanti, J. L. *Appl. Environ. Microbiol.* **72**, 6865–6875 (2006).
9. Limmathurotsakul, D. et al. *PLoS Negl. Trop. Dis.* **7**, e2072 (2013).
10. Dance, D. A. B. *Clin. Infect. Dis.* **60**, 251–253 (2015).

Lipid code for membrane recycling

The sequential action of enzymes has been shown to modify members of a class of membrane lipid called phosphoinositides to direct integral membrane proteins for recycling. [SEE LETTER P.408](#)

TAMAS BALLA

The ability of cells to take up nutrients or receive signals from their neighbours depends on proteins that are integrated into the plasma membrane. These proteins are not static, and can be moved to the cell's interior through a process called endocytosis. Some will subsequently be routed back to the cell surface and reused, and others will be degraded. There are grave consequences if the balance between recycling and degradation is derailed. In this issue, Ketel *et al.*¹ (page 408) define a molecular mechanism that underpins the decision to direct cargo for recycling.

Phosphatidylinositol (PI), a major class of phospholipid, is an integral component of cell membranes. Phosphorylation at one or more of three different positions (dubbed 3, 4 and 5) in the headgroup of PI, which protrudes from the membrane, can give rise to seven minor lipid species called phosphoinositides (PIPs). There is considerable evidence² to suggest that

PIPs are crucial intracellular traffic controllers. Different species of PIP recruit specific effector proteins, many of which direct the routing of membranes and their cargoes within the cell.

PI3P (a PIP marked by one phosphate group at position 3) is needed for the formation of membranous structures called early endosomes³, which contain endocytosed proteins. As early endosomes mature, their cargoes are sorted according to whether they will be recycled or degraded. Addition of a phosphate group to PI3P to form PI(3,5)P₂ directs proteins for degradation⁴. However, it has not been known which PIPs, if any, mark membranes for recycling.

A PIP phosphatase enzyme called MTM1 removes the phosphate from position 3 of PI3P and PI(3,5)P₂ (ref. 5). Mutations in MTM1 cause a severe, debilitating muscular disease in humans called X-linked centronuclear myopathy⁵. Ketel *et al.* analysed connective-tissue cells called fibroblasts taken from patients with X-linked centronuclear myopathy, and a

cultured cell line in which MTM1 was down-regulated by genetic manipulations. They observed that transferrin-receptor proteins, which are responsible for iron uptake and are normally recycled after endocytosis, became trapped in endosomes, unable to fuse with the plasma membrane in these cells.

Another protein that is normally recycled, β 1 integrin, also became trapped in endosomes. This observation is in agreement with a previous study, in which the *mtm* gene was mutated in flies⁶, and is interesting because the presence of β 1 integrin at the cell surface is crucial for many cell types, including muscle cells, to attach to their surrounding matrix⁷. Ketel *et al.* showed that it is the inability of the mutant MTM1 to remove phosphate from endosomal PI3P that prevents proper recycling of cell-surface molecules to the plasma membrane. This inability may contribute to the muscle defects seen in patients.

Next, Ketel *et al.* showed that, in parallel with the removal of phosphate from PI3P, a phosphate group is added to position 4 by a lipid kinase enzyme called PI4K2A, generating PI4P. The authors found that MTM1 and PI4K2A form a complex, and that PI4K2A is required for the recruitment of MTM1 to endosomes. PI4K2A and MTM1 also form a complex with the exocyst, a set of proteins that is essential for the fusion of recycling endosomes with the plasma membrane⁸. Thus, the authors have mapped a series of biochemical steps in membrane-protein recycling (Fig. 1) — beginning with the removal from the endosomal membrane of PI3P and its replacement with PI4P, and followed by the recruitment of various proteins — that prepare endosomes and their cargo proteins for return to the plasma membrane.

These results raise questions about the organization of the sorting process. Previous studies have suggested that PI4K2A is required for the degradation of some proteins (other receptor proteins such as epidermal growth factor receptors⁹, for example) and foreign particles ingested by immune cells called phagocytes¹⁰. PI4P is also found in late endosomes, which are intermediates for the degradation pathway¹¹. The ability to discriminate between the same PI4P signals in membrane compartments destined for different fates probably involves extra molecular regulators.

It has previously been postulated that PIP effectors function by recognizing additional membrane components, such as the Rab proteins¹². Rab11 is known to mark membranes destined for recycling, and Rab7 marks them for degradation¹². These proteins, together with PI4P, might recruit different molecular machines that determine different membrane fates. Similar processes are seen in other PIP-regulated pathways¹³, such as membrane-curvature sensing or the transport of certain lipids between the membranes of different

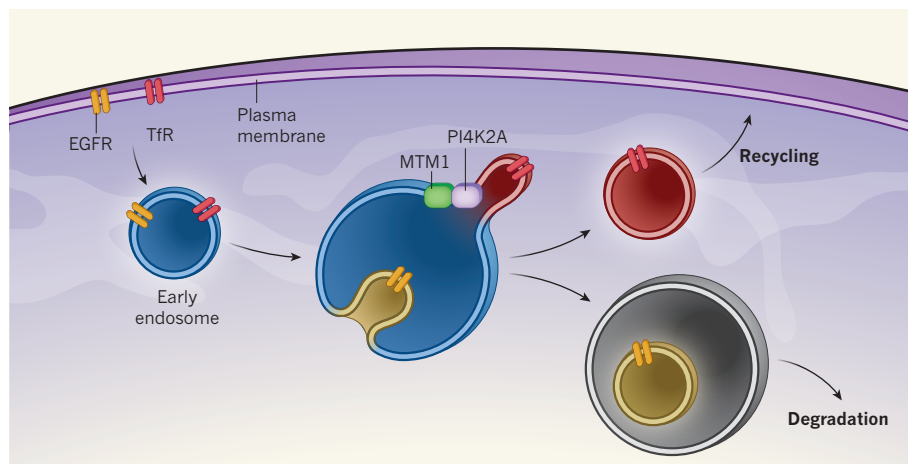


Figure 1 | Phospholipids direct membrane sorting. The proteins epidermal growth factor receptor (EGFR) and transferrin receptor (TfR) are removed from the cell surface in membranous structures called endosomes. The membranes of early endosomes contain PI3P (PI3P-containing membranes are indicated in blue) — a phospholipid of the phosphoinositide family that is tagged with a phosphate group at 'position 3'. As they mature, endosomes are sorted to determine the fate of the proteins they contain. In membrane regions such as those harbouring EGFR, PI3P is converted to PI(3,5)P₂ (yellow membranes), marking the region for degradation (the lipids present in the outer membrane of endosomes destined for degradation remain unknown). Ketel *et al.*¹ report that regions destined for recycling, such as those harbouring TfR, are modified by the sequential action of two enzymes. First, MTM1 removes the phosphate group from PI3P, then PI4K2A adds a phosphate group to position 4, generating PI4P (red membranes). The presence of PI4P directs the endosome back to the plasma membrane so that the proteins can be reused.

organelles, and is likely to be replicated in other lipid–protein complexes at different stages of endocytosis and recycling.

Ketel and colleagues' study marks a step forward in our understanding of how PIPs act as a 'lipid code' to direct the sorting of cell-surface molecules in endosomes. Moreover, they provide a neat demonstration of the importance of this code to preventing diseases such as X-linked centronuclear myopathy. Getting a grip on the molecular underpinnings of this debilitating disease may help scientists to devise strategies to improve its outcomes. ■

Tamas Balla is in the Program for Developmental Neuroscience, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA.
e-mail: ballat@mail.nih.gov

ECOLOGY

Biodiversity and productivity entwined

A systems-level analysis of grasslands across the planet provides stimulating insight into the interlaced pathways that connect species diversity and biological productivity in ecological communities. [SEE LETTER P.390](#)

KEVIN GROSS

An abiding goal of ecology is to understand how the biological diversity of natural communities is linked to the rate at which those communities produce biomass, if it is linked at all^{1–5}. Elucidating this connection and the processes behind it is useful, among other reasons, for anticipating how ecological communities may change in response to anthropogenic perturbations, such as alterations to the abiotic environment, the introduction of new species or the loss of established ones. Yet despite more than 40 years of study and debate, a comprehensive, mechanistic understanding of the relationship between biodiversity and ecosystem productivity remains elusive. On page 390 of this issue, Grace *et al.*⁶ argue that the interplay between these factors is more fully understood when both are placed in a rich network of cause-and-effect pathways, as opposed to being regarded as entities engaged in an isolated back-and-forth.

The authors considered data generated by the Nutrient Network⁷, a global scientific cooperative that examined several dozen grass-dominated plant communities from across the planet. Grace and colleagues analysed these data using a structural-equation model — a statistical method that examines

1. Ketel, K. *et al.* *Nature* **529**, 408–412 (2016).
2. Balla, T. *Physiol. Rev.* **93**, 1019–1137 (2013).
3. Simonsen, A. *et al.* *Nature* **394**, 494–498 (1998).
4. Takasuga, S. & Sasaki, T. *J. Biochem.* **154**, 211–218 (2013).
5. Amoasii, L., Hnia, K. & Laporte, J. *Curr. Top. Microbiol. Immunol.* **362**, 209–233 (2012).
6. Ribeiro, I., Yuan, L., Tanentzapf, G., Dowling, J. J. & Kiger, A. *PLoS Genet.* **7**, e1001295 (2011).
7. Volk, T., Fessler, L. I. & Fessler, J. H. *Cell* **63**, 525–536 (1990).
8. Wu, B. & Guo, W. *J. Cell Sci.* **128**, 2957–2964 (2015).
9. Minogue, S. *et al.* *J. Cell Sci.* **119**, 571–581 (2006).
10. Jeschke, A. *et al.* *Proc. Natl Acad. Sci. USA* **112**, 4636–4641 (2015).
11. Hammond, G. R. V., Machner, M. P. & Balla, T. *J. Cell Biol.* **205**, 113–126 (2014).
12. Wandinger-Ness, A. & Zerial, M. *Cold Spring Harb. Perspect. Biol.* **6**, a022616 (2014).
13. van Weering, J. R., Verkade, P. & Cullen, P. *J. Semin. Cell Dev. Biol.* **21**, 371–380 (2010).

This article was published online on 13 January 2016.

whether a hypothesized system of cause-and-effect pathways is consistent with covariation in a collection of variables. The starting point for this analysis was a putative network of cause-and-effect relationships for several aspects of ecosystems and their associated environments, devised from the authors' synthesis of the literature. The confrontation between data and model then occurred at two nested levels: between the sites of the Nutrient Network, and across replicate communities within each site.

Several provocative results emerge (Fig. 1). First, the authors find that the rate of biomass production increases with the number of species found at a site (its species richness). What is notable here is that this effect holds steady across the observed variation in species richness, instead of saturating in communities with greater richness, as a generation of experimental and theoretical work has suggested that it may⁸. Second, both between and within sites, a greater stock of accumulated above-ground biomass (live plant tissue and dead litter) decreases species richness. The authors suggest that this provides evidence that competition between species — primarily for light, at least at smaller spatial scales — is an important force in determining why communities contain as many species as they do. Again, the



50 Years Ago

It is particularly rewarding to investigate the experience of chronosystole, or time contraction during sympathetic excitation, as this state, which is connected with an increase in metabolic rate, can be elicited at will through psychotomimetic, pyretogenic drugs such as mescaline, marihuana, D-lysergic acid diethylamide (LSD), 'Psilocybin' ... No effort has been made ... to quantitatively differentiate the degrees of reactivity in subjects who have ingested the same dose of 'Psilocybin' ... We have refined the use of two already existing variables for the measurement of time contraction — handwriting samples and finger-tapping rates.

From *Nature* 22 January 1966

100 Years Ago

To discover whether the various objects carried about on the spines of the purple-tipped sea-urchin (*Echinus miliaris*) were accidentally picked up or deliberately placed there has recently formed the subject of a series of experiments by Mr. H. N. Milligan. He gives the results of his inquiry in the *Zoologist* for December. While usually stones, sea-weed, or shells are carried, tube-worms, hydroids, periwinkles, or tunicates, as chance may determine, are also used, apparently for the purposes of disguise and protection from enemies. That such objects are borne with a purpose, and not as a result of accident, is shown by the fact that when all foreign bodies are removed from the spines of urchins living in an aquarium, they will invariably be speedily replaced as soon as their loss is perceived, the tube-feet being used to perform this office. Young individuals were more assiduous in this regard than adults, but in all cases particular care was taken to conceal the anus, which is apparently a very vulnerable spot.

From *Nature* 20 January 1916

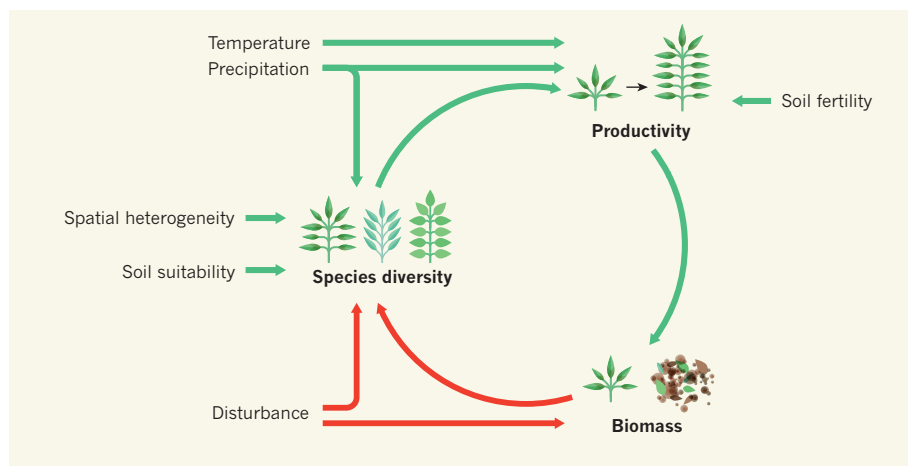


Figure 1 | A network-based view of biodiversity and biomass production. Using a structural-equation model and data from 39 herbaceous-plant communities across five continents, Grace *et al.*⁶ show that variation in biological diversity and biomass productivity is well explained by a network of cause-and-effect pathways. Positive pathways (green) have an enhancing effect, such as soil fertility increasing plant productivity; an example of a negative pathway (red) is the suppressive effect of above-ground biomass (live plant material and dead litter) on species diversity. The network depicted represents patterns that are consistent with variation between the 39 communities; variation within the communities supports a comparable network.

noteworthy aspect is that competition seems to be influential regardless of how productive a community is, contrary to the historical suggestion that competition is a stronger structuring force in more-productive communities^{1,5}. Finally, the analysis identifies several pathways by which the abiotic environment, including temperature, precipitation, soil properties and the history of major anthropogenic influences, adjusts the ecological context in which species diversity and productivity interact.

Herbaceous-plant communities provide the empirical grist for much of what we know about the relationship between biodiversity and productivity^{1–3,5}, for the good reason that these are systems in which those two entities can be readily quantified (relatively speaking, of course!). However, it is natural to ask how the causal network posited by Grace *et al.* can apply to other major biomes, such as tropical forests, coral reefs or the human body's resident microorganisms. A more explicit integration of the authors' causal network with the mathematical theories of species coexistence and community dynamics would help to delineate the network's biological generality.

Also, although a structural-equation model can reveal whether data are consistent with a causal network, it is possible that other causal networks are consistent with the data as well⁹. The authors justify their hypothesized network extensively, but alternative readings of the literature could surely be made. Whether other ecologically plausible causal networks can be constructed that are consistent with these or comparable data remains an open question, and one that should be asked.

Perhaps Grace and colleagues' most stimulating contribution, however, is their forceful suggestion that ecologists' theoretical,

experimental and statistical practices are sufficiently mature to allow a more holistic, systems-level view of the relationship between productivity and biodiversity. This suggestion challenges the decades-old view that the biodiversity–productivity relationship can be characterized by a simple yet general bivariate curve^{1,5}. This bivariate model has been remarkably durable, despite the severe and widely recognized weakness of its predictive power and disagreement regarding the basic shape of any such curve^{1–3,5}. Mounting dissatisfaction with the bivariate model has inspired several alternative, more-nuanced models

for how biodiversity and productivity are intertwined^{4,10,11}. To this collection, Grace *et al.* have contributed a framework that is unique in its deep integration with data gathered from around the globe.

Nevertheless, the staying power of the bivariate relationship testifies to its appeal as a first handle on the daunting complexity of complicated ecosystems. As the dust settles, ecologists will ultimately gravitate to the conceptual framework that provides the best balance of transparency, predictive capacity and fidelity to data. The efforts of Grace *et al.* and others set the stage for a rigorous investigation and evaluation of more-nuanced conceptual models of ecological communities. In time, perhaps these models will emerge alongside the bivariate diversity–productivity model as cornerstones of ecological thought and inquiry, or may even supersede it. ■

Kevin Gross is in the Biomathematics Program, North Carolina State University, Raleigh, North Carolina 27695, USA.
e-mail: kevin_gross@ncsu.edu

1. Grime, J. P. *Nature* **242**, 344–347 (1973).
2. Mittelbach, G. G. *et al. Ecology* **82**, 2381–2396 (2001).
3. Adler, P. B. *et al. Science* **333**, 1750–1753 (2011).
4. Willig, M. R. *Science* **333**, 1709–1710 (2011).
5. Fraser, L. H. *et al. Science* **349**, 302–305 (2015).
6. Grace, J. B. *et al. Ecology* **97**, 390–393 (2016).
7. Borer, E. T. *et al. Methods Ecol. Evol.* **5**, 65–73 (2014).
8. Cardinale, B. J. *et al. Nature* **486**, 59–67 (2012).
9. Kline, R. B. *Principles and Practice of Structural Equation Modeling* 4th edn (Guilford, 2015).
10. Schmid, B. *Trends Ecol. Evol.* **17**, 113–114 (2002).
11. Cardinale, B. J., Bennett, D. M., Nelson, C. E. & Gross, K. *Ecology* **90**, 1227–1241 (2009).

This article was published online on 13 January 2016.

EXOTIC ATOMS

Antimatter may matter

The charge neutrality of the antimatter atom antihydrogen has been confirmed with unprecedented accuracy, paving the way for experiments that could simultaneously solve several of physics' biggest mysteries. SEE LETTER P.373

THOMAS J. PHILLIPS

Something is wrong with our knowledge of antimatter. Our very existence proves this because, if our current understanding of antimatter were correct, the vast majority of the matter from which we are made would have been annihilated with antimatter shortly after the Big Bang. The problem is that we do not know what part of our understanding is wrong, so studies are needed of the basic properties of antimatter. On page 373 of this issue, Ahmadi *et al.* report a limit on the net

charge of antihydrogen — atoms comprised of antiprotons and positrons, the antimatter counterparts of protons and electrons, respectively. Although the researchers have not yet identified the difference between matter and antimatter that might explain our matter-dominated Universe, their experiment paves the way for measurements that could fundamentally change our understanding of the Universe.

Working with antimatter presents unique challenges because, as Star Trek fans know, antimatter annihilates on contact with matter. This means that any antimatter created on

CERN

Earth — such as positrons created through the radioactive decay of a nucleus, or antiprotons made using an accelerator — must be kept in a vacuum. Charged antiparticles such as these are easily manipulated by magnetic and electric fields, and sophisticated techniques have been developed for trapping and cooling them. Antihydrogen is created by mixing trapped antiprotons and positrons. However, because it is electrically neutral, once antihydrogen forms, it is no longer confined by the fields that trap the antiprotons and positrons.

Ahmadi *et al.* used an apparatus at CERN near Geneva, Switzerland (Fig. 1), that traps antihydrogen atoms using a magnetic-field minimum at the centre. The trap is very shallow, so it captures only the coldest (slowest) atoms. These cold antihydrogens are so rare that, on average, just one is trapped per trial; only a few trials can be run each hour. Fortunately, single antihydrogens can be counted with high efficiency because each antimatter atom annihilates to form several energetic particles when it escapes from the trap. The downside of this detection method is that, in the words of Joni Mitchell's *Big Yellow Taxi*, “you don't know what you've got till it's gone” — it was only when the researchers turned the trap off at the end of each trial that they knew whether they had caught any antihydrogen.

To check the neutrality of antihydrogen, the researchers applied varying electric fields to the trapped atoms (a method called stochastic acceleration). If the net charge of antihydrogen exceeds one-billionth of the charge of an electron, then the applied fields would heat the trapped antimatter atoms enough that a statistically significant fraction of them would be lost from the trap. No such loss was observed, so Ahmadi *et al.* were able to impose a limit to the charge of antihydrogen that is 20 times lower than their previously reported best limit², which itself was a million times lower than the best limit before that³. When combined with the best measurement of the charge of an antiproton⁴, the uncertainty on the measured charge of the positron is reduced by a factor of 25.

To be honest, no one expects to find a non-zero charge for antihydrogen — if we did find it, it would enormously complicate the theoretical picture, rather than pointing us towards a solution to the problem of the ‘missing’ antimatter. However, we need to know that antihydrogen is neutral before we can measure the gravitational force on it, because electromagnetic forces are so much stronger than gravitational forces. And, unlike the charge measurement, the gravity measurement could potentially solve not just the antimatter problem, but several other major problems, too.

In addition to the lack of antimatter, every list of the biggest mysteries in physics includes the natures of three things: dark energy, dark matter and cosmic inflation. These three were

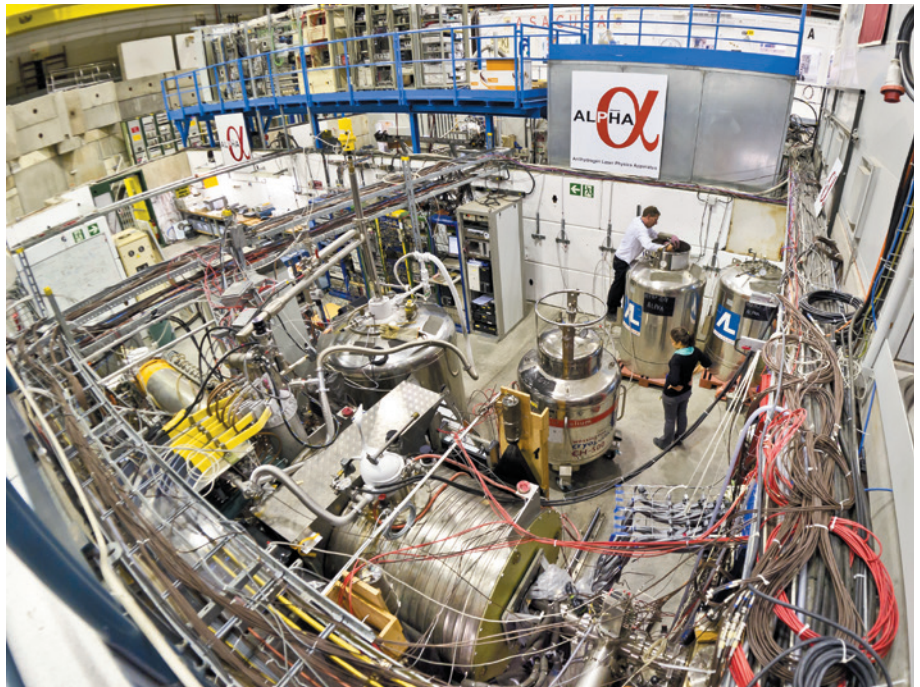


Figure 1 | The ALPHA facility at CERN near Geneva, Switzerland. Ahmadi *et al.*¹ have used the ALPHA facility to constrain the limit on the electrical charge of antihydrogen atoms.

all posited as ad hoc solutions to problems posed by cosmological observations that do not fit predictions arising from the general theory of relativity. Dark energy is needed to explain why the cosmic expansion is not slowing down; dark matter is invoked to resolve why galaxies are rotating too fast to be bound by gravity due to visible matter; and cosmic inflation is needed to explain how all parts of the Universe are the same temperature when the Big Bang occurred too quickly for everything to be causally connected (that is, regions of the visible Universe separated before light, and therefore temperature information, from one region could reach all other regions).

But all of these problems and more can be explained by a single change to our theory of gravity, a change that should soon be tested in antimatter experiments^{5–8}. The Dirac–Milne cosmological model⁹ is based on the assumption that the Universe has equal amounts of matter and antimatter that repel each other gravitationally, contrary to the prediction of general relativity¹⁰. Currently, all the observed galaxies in the Universe are thought to be formed of matter, because galaxies have been seen to collide but not to annihilate. But if matter and antimatter repel each other, they would never collide, so the missing antimatter could be hiding in plain sight.

Furthermore, in this model the Universe would have zero net gravitational charge, so dark energy would not be needed to explain why its expansion is not slowing down. Such a coasting Universe also obviates the need for cosmic inflation, because its expansion would have been sufficiently slow for the entire visible Universe to have been causally

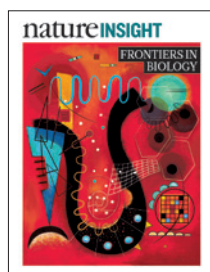
connected. Finally, antimatter that has negative gravitational charge would change the nature of vacuum (a process called vacuum polarization) such that it could provide enough gravity to hold galaxies together without the need for dark matter¹¹.

Occam's razor favours the antimatter-gravity solution to these problems, but it requires us to modify our theory of gravity, probably by expanding our understanding of the underlying geometry of space-time. Given that general relativity is fundamentally incompatible with quantum mechanics, perhaps we should not be so reluctant to consider this. Fortunately, experiments should soon tell us whether gravity is the problem with antimatter, and, if they do, then I predict that physicists will quickly find that this should have been obvious. Perhaps it already is. ■

Thomas J. Phillips is in the Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616, USA.
e-mail: phillips@phys.iit.edu

1. Ahmadi, M. *et al.* *Nature* **529**, 373–376 (2016).
2. Amole, C. *et al.* *Nature Commun.* **5**, 3955 (2014).
3. Greenland, P. T. *Contemp. Phys.* **38**, 181–203 (1997).
4. Hori, M. *et al.* *Nature* **475**, 484–488 (2011).
5. Doser, M. *et al.* *Class. Quantum Grav.* **29**, 184009 (2012).
6. Perez, P. & Sacquin, Y. *Class. Quantum Grav.* **29**, 184008 (2012).
7. The ALPHA Collaboration & Charman, A. E. *Nature Commun.* **4**, 1785 (2013).
8. Kaplan, D. M. *et al.* *EPJ Web Conf.* **95**, 05008 (2015).
9. Benoit-Lévy, A. & Chardin, G. *Astron. Astrophys.* **537**, A78 (2012).
10. Nieto, M. M. & Goldman, T. *Phys. Rep.* **205**, 221–281 (1991).
11. Hajdukovic, D. S. *Astrophys. Space Sci.* **334**, 215–218 (2011).

FRONTIERS IN
BIOLOGY



Cover illustration
'Cryptic' by
Nik Spencer

Editor, *Nature*
Philip Campbell

Publishing
Richard Hughes

Insights Editor
Ursula Weiss

Production Editor
Elizabeth Batty

Art Editor
Nik Spencer

Sponsorship
Reya Silao

Production
Ian Pope

Marketing
Steven Hurst

Editorial Assistant
Rebecca White

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel: +44 (0) 20 7833 4000
e: nature@nature.com



nature publishing group

The *Nature Insight* 'Frontiers in biology' aims to cover timely and important developments in biology, ranging from the subcellular to the organismal level, and including molecular mechanisms and biomedicine.

The collection begins with a Review on metastasis, which remains a significant challenge in the treatment of cancer. Metastases in some organs are less responsive to therapy than those in others. Joan Massagué and Anna Obenauf discuss the processes that underlie the colonization of various organs by cancer cells that escape from the primary tumour and enter the circulatory system. Improved understanding of these processes could facilitate the early diagnosis of metastases and lead to new therapeutic interventions.

Next, Michael Karin and Hans Clevers explore the cellular and molecular mechanisms of protective inflammation in tissue repair and regeneration, with a focus on the intestines and the liver. The positive effects of inflammation play an essential part in the restoration of tissue homeostasis, and offer protection from chronic inflammation, fibrosis and cancer.

Endothelial cells form the lining of blood vessels, but they also function in the development, growth and regeneration of organs through their release of paracrine signals known as angiocrine factors. Shahin Rafii and colleagues examine the emerging role of endothelial cells as signalling centres that orchestrate organ regeneration, morphogenesis and homeostasis.

Cells counteract endoplasmic reticulum stress — the abnormal accumulation of misfolded and unfolded proteins of the secretory pathway — by activating a homeostatic process called the unfolded protein response. Miao Wang and Randal Kaufman summarize our current understanding of this field at the molecular and cellular levels, as well as in relation to disease.

Resistance to antibiotics poses one of the most pressing threats to modern health care. Eric Brown and Gerard Wright consider the history and use of antibiotics, the factors that have led to the alarming rise in resistant pathogens, and the prospects for overcoming hurdles that hold back antibiotic discovery in the twenty-first century.

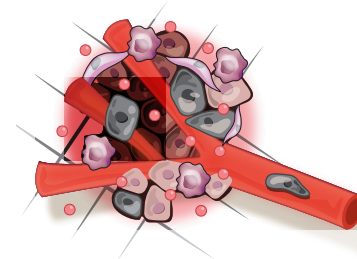
**Barbara Marte, Sadaf Shadan, Clare Thomas,
Christina Tobin Kährström & Ursula Weiss**
Senior Editors

CONTENTS

REVIEWS

298 Metastatic colonization by circulating tumour cells

Joan Massagué & Anna C. Obenauf

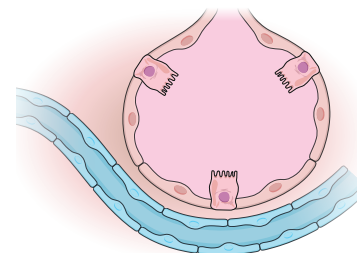


307 Reparative inflammation takes charge of tissue regeneration

Michael Karin & Hans Clevers

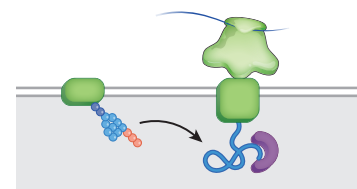
316 Angiocrine functions of organ-specific endothelial cells

Shahin Rafii, Jason M. Butler & Bi-Sen Ding



326 Protein misfolding in the endoplasmic reticulum as a conduit to human disease

Miao Wang & Randal J. Kaufman



336 Antibacterial drug discovery in the resistance era

Eric D. Brown & Gerard D. Wright

Metastatic colonization by circulating tumour cells

Joan Massagué¹ & Anna C. Obenauf[†]

Metastasis is the main cause of death in people with cancer. To colonize distant organs, circulating tumour cells must overcome many obstacles through mechanisms that we are only now starting to understand. These include infiltrating distant tissue, evading immune defences, adapting to supportive niches, surviving as latent tumour-initiating seeds and eventually breaking out to replace the host tissue. They make metastasis a highly inefficient process. However, once metastases have been established, current treatments frequently fail to provide durable responses. An improved understanding of the mechanistic determinants of such colonization is needed to better prevent and treat metastatic cancer.

Malignant tumours start early on the road to metastasis. Cancer cells that are invasive and motile can enter the circulatory system long before a tumour is diagnosed. Most of these cells will perish, but a small proportion manages to infiltrate distant organs and survives as disseminated seeds for eventual relapse. Thus, at diagnosis, a primary tumour might have already seeded distant organs with thousands of cancer cells. These cells will face many obstacles before they can take over their host organ and form clinically relevant lesions. Indeed, organ colonization is the most complex and rate-limiting phase of metastasis.

Research on metastatic colonization has been hindered by the complexity of the biology and a lack of adequate experimental models. However, the development of patient-derived and genetically engineered mouse models of metastasis, enhanced imaging technologies and advanced genomic sequencing, including the ability to analyse single cells, together with improved access to clinically annotated tissue samples, has brought fresh insight to our understanding of the molecular mechanisms that allow circulating tumour cells (CTCs) to invade distant organs, settle in supportive niches and eventually overtake their host tissue (Fig. 1). This has allowed us to better conceptualize the metastatic process as a whole and provides a basis for superior treatments. Such advances show that the elucidation of metastatic colonization is a tractable problem with clinical benefits. In this Review, we highlight current concepts and open questions at the forefront of this field. Each aspect of metastatic colonization that we explore is an individually rich area of research, and we cite specialized articles that cover them in depth.

The inefficiency of metastatic colonization

Even small tumours can shed millions of cancer cells. However, many people who have recovered from cancer never experience relapse or do so only after a long period of latency without clinically manifest disease. The number of CTCs — cancer cells that are found in blood from patients — far exceeds the number of overt metastatic lesions that develop¹. Cancer cells that survive after infiltrating distant organs, known as disseminated tumour cells (DTCs), can be present in the bone marrow of people with cancer for years, yet only about half of these patients develop overt metastasis². These clinical observations suggest that metastatic colonization is a very inefficient process in which most cells die and only a minority of those that survive form macrometastases.

Data from experimental mouse models are in line with the clinical evidence. For example, intravenously injected cancer cells that reach the

lungs die in large numbers within 2 days³, as do arterially injected cancer cells that lodge in the brain, liver or bone marrow⁴. And populations of cells that are enriched for highly metastatic cells experience extensive attrition after infiltration of distant organs⁵. The vast majority of melanoma cells that are injected in the portal vein fail to form micrometastases in the liver, and only 0.02% form macrometastases^{6,7}. Similarly, most cancer cells that infiltrate the brain will die^{8–10}. Such inefficiencies cannot be attributed simply to a scarcity of cancer stem cells with metastasis-initiating potential: the majority of breast cancer stem cells that reach the lungs undergo apoptosis¹¹ and colorectal cancer stem cells are cleared quickly after infiltrating the liver parenchyma¹². Observations such as these, in mouse models and in the clinic, imply that the factors that influence the survival and tumour-initiating activity of DTCs are important determinants of metastasis.

Organ infiltration

The early steps in the metastatic cascade, which include the invasion and migration of cancer cells into tissues and the circulatory system, have been extensively studied^{13,14}. Cytoskeletal rearrangements within cancer cells¹⁵, combined with adhesive interactions between cells and the secretion of extracellular-matrix-degrading metalloproteinases (MMPs) and cathepsins^{16,17}, drive their invasion and migration through the stroma, a network of supportive, connective tissue cells. Cancer cells can migrate as single cells boring a path through the extracellular matrix¹⁸, move along collagen fibres¹⁹ or migrate collectively as ensembles that forge ahead of the tumour invasion front²⁰. In prostate cancer, invasion along nerve fibres provides an additional route for dissemination²¹. In response to transforming growth factor- β (TGF- β) and other cell-signalling proteins that are released by stromal cells, carcinoma cells can undergo epithelial-to-mesenchymal transition (EMT) — a reversible phenotypic change in which cells lose intercellular adhesion and epithelial polarization and gain motility and invasiveness²². EMT has an important role in gastrulation and other morphogenic events that occur during development. In carcinoma cells, EMT can promote cell entry into the vasculature, known as intravasation, and support the induction of a stem-cell phenotype, whereas a reversal of this state after extravasation — in which cancer cells exit capillaries to enter organs — can facilitate colonization²³. However, studies in models of breast and pancreatic cancer suggest that EMT could be dispensable for the establishment of metastasis, despite it contributing to the aggressiveness of cancer cells by increasing their

¹Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, New York 10065, USA. [†]Present address: Research Institute of Molecular Pathology, Vienna Biocenter, 1030 Vienna, Austria.

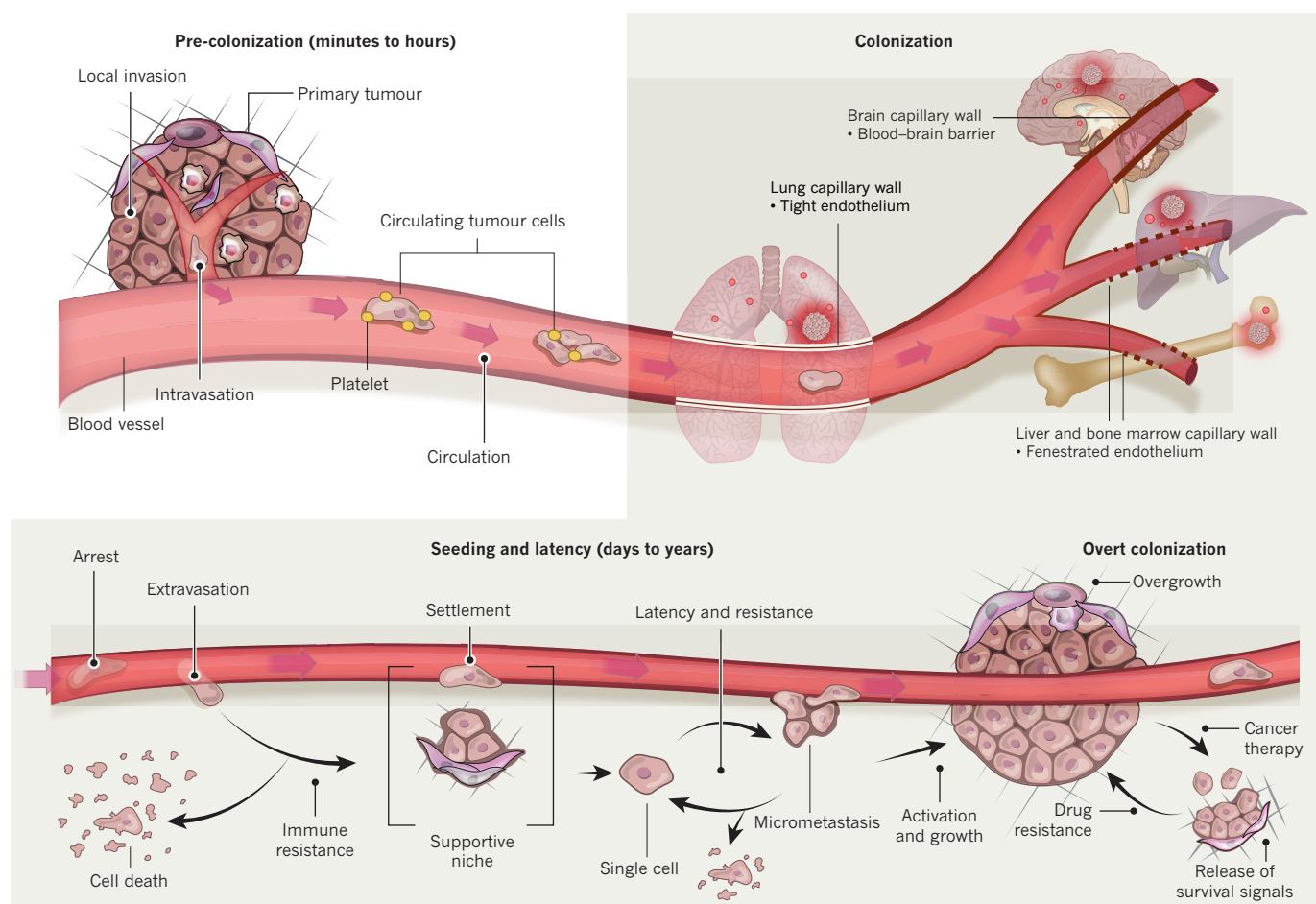


Figure 1 | Metastatic colonization. Metastasis proceeds through multiple steps and restrictive bottlenecks. The pre-colonization phase of metastasis comprises a series of events that occur on a timescale of minutes to hours. Local invasion of the primary tumour by cancer cells is followed by their intravasation into the tumour vasculature. The cancer cells then enter the circulatory system as single cells or clusters that are coated with platelets. Circulatory patterns, which move blood through the lungs and then on to other organs, and the differing structure of the capillary walls in each organ influence the dissemination of CTCs. On their arrest in capillaries at distant sites, the cancer cells extravasate into the parenchyma of target organs to commence colonization. Colonization can be parsed into many steps that occur on a timescale of years. After extravasation, colonizing cancer cells must develop resistance to immunity and other host-tissue defences to survive. Settlement in

supportive niches enables them to survive and retain their stem-like tumour-initiating capacity. The cancer cells then enter a latent state as single cells or micrometastases. During latency, which can last from months to decades, disseminated cells must achieve long-term survival. They might also acquire traits that are required to overtake host tissue. When the cancer cells break out of latency, they reinitiate overt outgrowth and overtake the local tissue microenvironment. Therapeutic treatment can partially eliminate clinically manifest metastases. However, under therapy-induced stress, cancer cells and non-neoplastic stromal cells mobilize survival signals that nurture the residual disease until minority drug-resistant clones emerge to lead the outgrowth of a drug-resistant tumour. Different host-tissue microenvironments select for cancer cells with distinct metastatic traits, which gives rise to organ-specific populations of metastatic cells.

chemoresistance^{24,25}. Therefore, the contribution of EMT to metastasis might be more nuanced than thought previously.

Cancer cells can leave tumours as single cells or in clusters (Fig. 1). There is growing evidence to indicate that distinct cancer-cell clones show cooperative behaviour, which promotes their mutual survival and metastatic ability^{26–29}. Polyclonal metastatic seeding, for example, has been documented in patients with prostate cancer³⁰, and in experimental models, polyclonal clusters of CTCs establish metastases more efficiently than single cells³¹.

In the bloodstream, cancer cells are exposed to considerable shear forces, the innate immune system and oxidative stress. To protect themselves during transit, they associate with platelets³² and undergo reversible metabolic changes that increase their ability to withstand oxidative stress³³. For instance, melanoma cells show an increased dependence on nicotinamide adenine dinucleotide phosphate (NADPH)-generating enzymes in the folate pathway, and inhibition of this pathway reduces overt metastasis³⁴.

CTCs become mechanically entrapped in capillaries before they exit the blood and move into tissue. This is considered to be the main mechanism

for cancer-cell arrest. The first capillary bed that a CTC encounters is determined by patterns of blood circulation in the body. In most organs, the venous circulation leads to the right ventricle of the heart and on to the lungs, although the venous circulation from the gut drains into the liver. The resulting retention of CTCs in the lungs or liver, respectively, contributes to the high incidence of metastasis in these organs³⁵. However, some CTCs bypass these initial filters, perhaps through larger arteriovenous shunts, to reach other organs through the arterial circulation.

CTCs that lodge in the microvasculature can initiate growth within the lumen to form an embolus that eventually ruptures the vessel³⁶ or extravasate by breaching vascular walls^{5,6}. The composition of these walls differs between organs and is another factor that influences where cancer cells extravasate (Fig. 1). The capillaries in the liver and bone marrow, called sinusoids, are lined with fenestrated endothelial cells and a discontinuous basal lamina³⁷ — gaps that might facilitate the extravasation of CTCs and contribute to the high incidence of liver and bone metastasis^{13,38}. In contrast, the endothelium of lung capillaries has tight junctions and a basement membrane, and the capillary walls of the brain are reinforced by pericytes and the processes of astrocytes, which together constitute

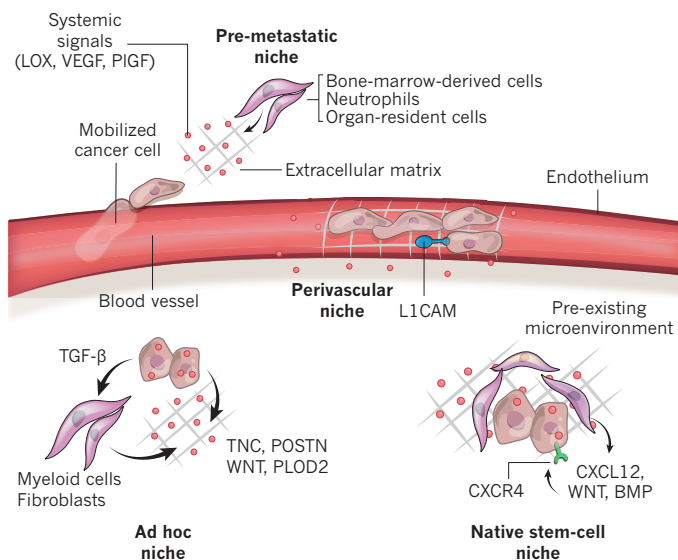


Figure 2 | Metastatic niches. Cancer cells that infiltrate distant tissues survive and retain their stem-cell potential by locating themselves in supportive niches, which are akin to the niches that support normal adult stem cells. A number of different niches have been proposed. Pre-metastatic niches form before the arrival of cancer cells by systemic signals from the primary tumour that recruit supportive stromal cells. Perivascular niches support cancer cells that spread over the capillary basement membrane after extravasation. The cancer cells remain close to cells of the endothelium and their paracrine factors. Ad hoc niches are established by the secretory products of cancer cells and act in an autocrine manner or recruit stromal components as sources of supportive signals. Native stem-cell niches of the host tissue are often invaded by the infiltrating cancer cells, which allows the cells to occupy directly a supportive microenvironment. The location or composition of these niches can overlap. For example, a native stem-cell niche might be in a perivascular site, or pre-metastatic signals might combine with those released by the cancer cell at an ad hoc niche.

the blood–brain barrier^{13,37}. Diverse genes have been identified in model systems that mediate the extravasation of breast-cancer CTCs in the lungs and are also associated with lung metastasis in the clinic. These include the protein Fascin-1 and other components of invading protrusions known as invadopodia³⁹, autocrine enhancers of cancer-cell motility such as ephrins and WNT ligands^{39,40}, and mediators of endothelial disjunction and vascular permeability that include angiopoietin-like 4 (ANGPTL4), vascular endothelial growth factor (VEGF), the cyclooxygenase COX2 (also known as prostaglandin-endoperoxide synthase), MMP1 and osteonectin^{41–44}. Platelets that are associated with CTCs can stimulate extravasation by releasing TGF- β and triggering EMT in the cancer cells³² or by secreting adenine nucleotides, which relax endothelial cell junctions⁴⁵. Physical contacts between macrophages and CTCs help to pull CTCs across capillary walls in the lungs⁴⁶. Many of these factors also enhance the extravasation of CTCs in the brain, which involves the cancer-cell-derived sialyltransferase enzyme ST6GALNAC5 (ref. 47), cathepsin S⁴⁸ and microRNAs mir-105 and mir-181c^{49,50}. These mediators each provide a finite increase in the probability of metastatic seeding and frequently act in parallel. In sum, a combination of priming signals from the tumour stroma, the composition of CTC clusters, blood-circulation patterns, the structure of target-organ capillary walls and cancer-cell-autonomous functions determine metastatic infiltration of specific organs.

Tissue defences against infiltrating cancer cells

In primary tumours, cancer cells develop in a co-evolving microenvironment that suppresses immune surveillance^{17,51}. However, because support is not available immediately to cancer cells as they infiltrate distant organs, most of these cells will die^{5–7}. The ‘seed and soil’ hypothesis was based on the observation that different cancers show a predilection for metastasis in different organs and predicted that certain organs are more hospitable

to wandering cancer cells than others⁵². Seed and soil is an appealing metaphor but can be misleading. To disseminating cancer cells, every distant soil is deadly, although some soils can be less deadly than others. In fact, the most welcoming of all soils could be the primary tumour itself. The preferential reseeding of primary or metastatic tumours by CTCs, rather than the seeding of tumour-free secondary sites, is called tumour self-seeding⁵³. In experimental systems, self-seeding can amplify the most aggressive cancer-cell clones⁵³ and disperse drug-resistant clones during the treatment of metastatic melanoma with targeted therapy⁵⁴.

In a new and challenging microenvironment, freshly disseminated cancer cells can be particularly vulnerable to immune surveillance (Fig. 1). Major players in antimetastatic immune surveillance include cytotoxic T cells and natural killer (NK) cells⁵⁵. For instance, the depletion of cytotoxic T cells or NK cells has been shown to increase metastasis^{56,57} and inhibition of the tyrosine kinase Mer, a negative regulator of NK cells, suppresses metastasis⁵⁸. Moreover, the specific immune-cell composition of an organ can influence the organ’s susceptibility to overt metastasis. The liver, for example, is rich in NK cells. Neutralization of pro-apoptotic NK-cell-derived tumour necrosis factor (TNF)-related apoptosis-inducing ligand (TRAIL), or the genetic depletion of NK cells, increases hepatic metastasis in mice^{59,60}. Advances in immunotherapy, most notably those that use immune checkpoint inhibitors, have yielded striking results in metastatic melanoma and other tumour types^{61,62}. Thus, immunity is a major defence against metastasis.

Other cell types can also mount a strong defence against metastatic infiltration. Astrocytes, the most abundant type of cell in the brain, reject extravasated cancer cells by releasing the serine protease plasminogen activator (PA). PA converts the zymogen plasminogen to plasmin, which then mobilizes the pro-apoptotic cytokine Fas ligand (FasL) to kill infiltrating cancer cells. To avert this fate, brain metastatic cells from breast and lung adenocarcinomas produce the PA inhibitors neuroserpin and serpin B2 (ref. 9).

Supportive niches

Adult stem cells reside in specialized niches that provide cues that help to maintain a balance between stem-cell proliferation and quiescence as well as self-renewal and differentiation. Stem-cell niches are rich in developmental and self-renewal signals, such as hedgehog, Wnt, members of the TGF- β family and the chemokine CXCL12 (refs 63, 64). Tumours are thought to arise from mutant stem cells in their native niches or from the progeny of cells that retain their tumour-initiating capacity and benefit from these niche signals^{65–68}. After cancer stem cells disperse to distant sites, their survival and tumour-initiating potential can benefit similarly from interactions with specialized niches⁶⁹ (Fig. 2). Evidence suggests that prostate-carcinoma stem cells occupy native haematopoietic stem-cell niches in the bone marrow⁷⁰. The CXCL12 receptor CXCR4 is a marker and mediator of breast-cancer metastasis to CXCL12-rich bone-marrow sites⁷¹. Breast tumours that are rich in a CXCL12-secreting mesenchymal stroma select for CXCL12-responsive cancer-cell populations that are predisposed to survive in the bone marrow⁷².

Small blood vessels are surrounded by a space that is rich in supportive signals and can favour cancer stem-cell growth and resistance to anticancer therapy^{73,74} (Fig. 2). A striking example of the interaction of metastatic cells with perivascular sites is observed in brain metastasis by breast-cancer, lung-cancer and melanoma cells in which the extravasated cancer cells remain closely associated with capillaries^{8,75}. The cells spread along the basal lamina that surrounds the capillaries and proliferate to form a sheath that eventually engulfs and remodels the co-opted capillary network — a process that is mediated by expression of the cell-adhesion molecule L1CAM in the metastatic cells⁹ (Fig. 2).

CTCs can set up an ad hoc niche by producing components of stem-cell niches themselves (Fig. 2). Lung-metastatic breast cancer cells produce the extracellular-matrix protein tenascin C, which is deposited in the incipient colony to amplify Notch and Wnt signalling⁷⁶. Breast-cancer stem cells can also secrete TGF- β , which stimulates stromal fibroblasts to produce

periostin, a binding partner of tenascin C that recruits Wnt factors¹¹. The secretion by cancer cells of the collagen-crosslinking enzymes LOX and PLOD2, which stiffen the extracellular matrix, amplifies integrin–focal adhesion signalling and also favours metastasis^{77–79}.

Experimental models have provided evidence that systemic signals from primary tumours can influence the microenvironment of distant organs by creating pre-metastatic niches before the arrival of CTCs^{80,81} (Fig. 2). Different classes of systemic mediators, such as tumour-derived inflammatory cytokines, exosomes and extracellular-matrix-remodelling enzymes, have been shown in breast-, lung- and gastrointestinal-tumour models to recruit bone-marrow-derived cells and precondition the lung, liver or bone marrow for infiltrating cancer cells^{82–85}. Melanoma cells secrete exosomes that might induce vascular leakiness and inflammation during the formation of pre-metastatic niches⁸⁶. Similarly, macrophage inhibitory factor (MIF)-containing exosomes that are released by pancreatic cancer cells increase liver metastasis by inducing TGF- β secretion, stimulating the production of the glycoprotein fibronectin in hepatic stellate cells and recruiting bone-marrow-derived cells to the liver⁸³. Integrins were proposed to target cancer-cell-derived exosomes to specific organs to unload their cargo and prepare the organ for the arrival of tumour cells⁸⁷.

Observations from the clinic raise questions about the role of pre-metastatic niches in people with cancer. Most patients develop metastasis months or years after the removal of their primary tumour, until which time tumour cells remain largely dormant. Yet in experimental models, pre-metastatic niches support the immediate outgrowth of DTCs. Research must address whether pre-metastatic niches remain primed for years after the removal of a primary tumour or, alternatively, if their role is to enhance the survival of infiltrating cells to increase their numbers before they enter a latent state.

Growth and survival pathways

A plethora of genes and signals support metastatic cell growth and survival in experimental models, and the expression of these genes can predict relapse in the clinic (Box 1). Many such pro-metastatic stromal mediators ultimately activate stem-cell support pathways, such as the Wnt, TGF- β , bone morphogenetic protein (BMP), Notch and signal transducer and activator of transcription (Stat)3 pathways. Others activate pathways

that integrate cell metabolism and survival, including the phosphoinositide 3-kinase–protein kinase B (PI3K–AKT), mitogen-activated protein kinase (MAPK) and hypoxia-inducible factor (HIF) pathways. Positional and mechanical pathways (Hedgehog and Hippo) and inflammatory pathways (NF- κ B and Stat1) are also activated⁶⁹. These pathways also drive development and tissue regeneration, but what is distinctive in the case of metastasis are the strategies that cancer cells employ to ensure that sufficient pathway activation is achieved in microenvironments with low levels of activating signals (Fig. 3). DTCs seem to be selected according to their ability to use whatever cues the host tissue offers.

By expressing autocrine pathway activators or by recruiting stromal cells to produce them (Fig. 3), DTCs can stimulate a crucial pathway or amplify the pathway's signalling output⁶⁹. For example, Stat3 stimulation by autocrine interleukin (IL)-6 mediates metastasis in prostate cancer cells, and PI3K–AKT stimulation by autocrine insulin-like growth factor 2 (IGF2) mediates metastasis in oesophageal cancer cells^{88,89}. The intracellular tyrosine kinase Src amplifies the ability of stromal CXCL12 to activate PI3K–AKT signalling in breast cancer cells that infiltrate the bone marrow⁹⁰. Breast cancer cells produce colony-stimulating factor 1 (CSF1) to recruit tumour-associated macrophages as a source of epidermal growth factor (EGF)⁹¹ or secrete CXCL1 to recruit myeloid precursors as a source of calprotectin (S100A8/9) for MAPK activation⁹². Colorectal cancer stem cells that reach the liver express TGF- β to recruit mesenchymal cells as a source of IL-11 for Stat3 activation in the cancer cells¹².

Cancer cells can also obtain vital support through physical contact with stromal cells (Fig. 3). Claudin-2-mediated cell–cell interactions between breast cancer cells and hepatocytes induce c-Met (hepatocyte growth factor receptor) signalling and stimulate metastasis to the liver⁹³. Contact between membrane vascular cell adhesion molecule 1 (VCAM1) on breast cancer cells that have infiltrated the lungs and α 4 integrins on stromal monocytes and macrophages activates PI3K–AKT signalling in the cancer cells⁹⁴. By contrast, contact between VCAM1 from breast cancer cells that exit dormancy in the bone marrow and α 4 integrins on monocytic precursors accelerates the differentiation of these precursors into osteoclasts that mediate osteolytic metastasis⁹⁵.

The activity of pro-metastatic pathways can be further increased by epigenetic alterations. For example, von Hippel–Lindau (VHL)-mutant renal-cell-carcinoma cells gain metastatic activity in multiple organs

BOX 1

Origin of metastatic traits

Metastasis develops through genetic and epigenetic changes and subsequent selection for favourable traits under the pressure of successive bottlenecks^{138,139}. Genomic comparisons show close clonal relationships between primary tumours and their metastases. Specific ancestors of metastatic clones can often be identified in the primary tumour^{30,139,140}, which supports the hypothesis that late clonal expansion in the primary tumour gives rise to metastasis-competent clones. These studies also provide evidence that metastases seed further metastases³⁰. Disseminated cancer cells remain dependent on the oncogenic mutations that underlie the primary tumour, which provides a basis for treating metastasis with drugs that target these oncogenic drivers¹²⁵. Oncogenic mutant alleles accumulate in metastases, such as the gain of mutant *KRAS* in pancreatic cancer metastasis¹⁴¹ and *TP53* and androgen-receptor mutations in prostate cancer metastasis¹⁴². However, no recurrent metastasis-specific mutations have been identified so far, which suggests that epigenetic alterations and other mechanisms of modifying gene expression are the predominant source of selectable pro-metastatic traits during clonal evolution in metastasis^{139,140}.

The cell-autonomous traits that favour the dissemination of cancer cells, their resistance in the circulation, extravasation and initial survival

in distant organs become important as soon as the cells leave the primary tumour, and are pre-selected in the primary tumour. For example, certain mediators of neoangiogenesis in breast tumours, such as COX2, epiregulin, MMP1 and VEGF, are repurposed by cancer cells for extravasation in the lungs and brain^{42,47,143}. Stromal TGF- β in triple-negative breast carcinomas — which do not express the oestrogen receptor, progesterone receptor or HER2 — induces the expression of the protein angiopoietin-like 4 (ANGPTL4) and primes the cancer cells for extravasation in the lungs⁴¹. These early metastatic traits can be selected under the stresses of tissue invasion, immune surveillance or hypoxia. The evidence favours a model in which a considerable proportion of cancer cells in a primary tumour acquire pro-metastatic traits that confer a finite probability of success in the early steps of metastasis. Clones with the most effective combination of pro-metastatic traits are most likely to give rise to metastatic lesions and also to re-seed the primary tumour. Beyond these early steps, cancer cells continue to evolve after their dissemination to distant organs to acquire traits for overt colonization, as demonstrated in bone metastasis¹¹⁴. Cancer cells that undergo early dissemination could evolve in parallel with, but independently from, the primary tumour¹⁴⁴. The origin of metastatic traits remains a fertile area for future research.

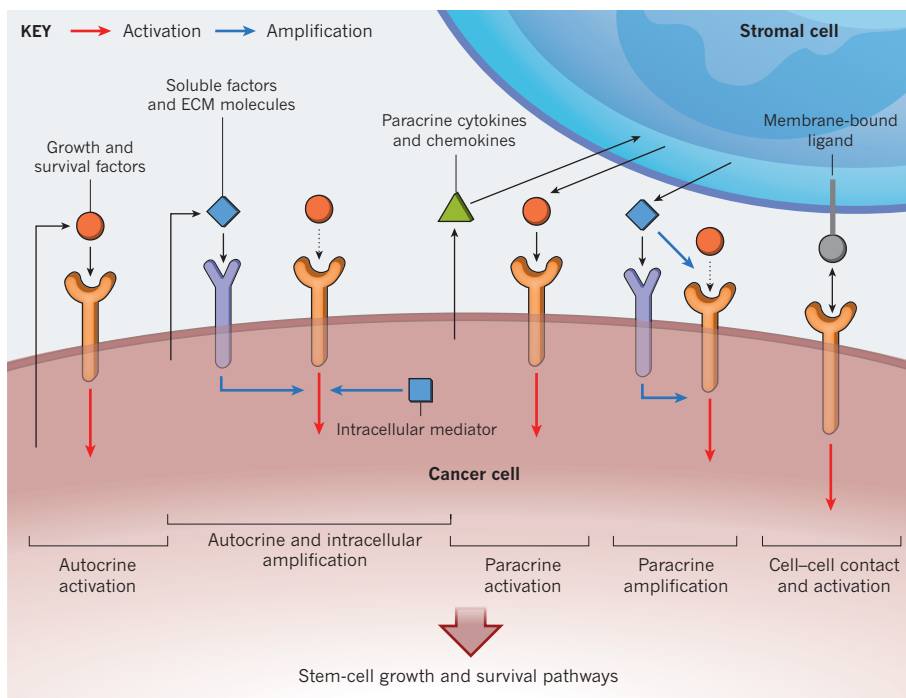


Figure 3 | The activation of growth and survival pathways by disseminated cancer cells. During colonization, metastasis-initiating cells require the activity of a common set of pathways that support the survival and growth of stem and progenitor cells. After infiltrating distant tissues, the disseminated cancer cells express paracrine mediators (such as IL-6 and IGF2), ECM components (such as TNC, LOX and PLOD2) and intracellular mediators (such as SRC and Ezrin) that activate and amplify survival and growth pathways. They also express paracrine factors (such as CSF1, CXCL1 and TGF- β) that help to recruit stromal cells as a further source of soluble activators and amplifiers that include the cytokines IL-6, WNT, EGF and S100A8/9 and the ECM components TNC and POSTN. Cancer cells can also achieve pathway activation through cell–cell contact using receptor–ligand pairings such as VCAM1– α 4 integrin and Jagged1–Notch.

through changes in DNA methylation and histone acetylation that expand the range of HIF target genes — the dominant oncogenic pathway in these cells⁹⁶. Additional inputs come from the expression of microRNAs that promote or suppress metastasis by regulating multiple mediators of tumour–stroma interactions^{97–100}. These examples show that DTCs resort to diverse mechanisms to procure vital inputs for survival and the retention of their tumour-initiating capacity.

Exactly when, where and how cancer cells turn to these various stromal cues is unclear. Are these niches and pathways important for all stages of metastatic colonization, at all times? Some might be crucial only after extravasation, when cancer cells are challenged by tissue defences or during the latent phase of metastasis in which cancer cells subsist for years without outgrowth. Yet others might be important only when DTCs exit dormancy and enter a proliferative state. Such questions remain unanswered because most experimental models of metastasis do not incorporate a latency phase. This gap in knowledge is also of concern from a translational perspective. Treating overt metastasis by targeting a survival mechanism that was relevant only during the initial infiltration of distant organs might have no clinical benefit. Likewise, targeting an oncogenic driver pathway in latent DTCs could be futile. But targeting mechanisms that support the viability of latent DTCs could actually eradicate residual disease.

Latent metastasis

The observation that patients relapse with metastatic disease months or years after removal of their primary tumour, combined with the detection of DTCs in the bone marrow of people with no evidence of metastatic disease, demonstrates that cancer cells that disseminate before treatment of the primary tumour retain the ability to initiate metastatic growth long thereafter. Some organs are more permissive than others to the accumulation of latent DTCs. For example, people with colorectal or gastric cancer can harbour DTCs in the bone marrow, yet the incidence of bone metastasis in these patients is low¹⁰¹. The incidence of DTCs in the bone marrow is a predictor of metastasis not only to bone, but also to the liver, lungs and brain².

Tumour dormancy is thought to occur through two modes (Fig. 1). In cellular dormancy, isolated DTCs enter a state of proliferative quiescence. Indeed, most DTCs in bone-marrow samples from people with cancer are found as quiescent single cells^{101–103}. But in tumour mass dormancy, micrometastases cease to grow because of insufficient vascularization or

constant culling by immune defences¹⁰². It is uncertain which mode most frequently leads to overt metastases.

Despite the biological and clinical relevance of metastatic latency, little is known about the mechanisms by which cancer cells enter a dormant state and the signals that sustain it, the niches that dormant cancer cells inhabit and what triggers the resumption of the cells' aggressive growth. The paucity of experimental model systems that incorporate a latent phase and the cost of studying such a process over an extended period in animal models have hindered progress. However, stromal signals that impose tumour dormancy have been identified in mouse xenograft models. TGF- β and BMPs, members of the TGF- β family, can enforce quiescence and inhibit self-renewal in carcinoma DTCs^{104–106}. The perivascular niche has also been implicated in the induction of cancer-cell dormancy¹⁰⁷. But environments that are rich in type I collagen¹⁰⁸ or fibronectin¹⁰⁹ inhibit dormancy.

A scarcity of stromal growth factors and an abundance of growth-inhibitory cues can favour metastatic dormancy in experimental models. On their own, however, these signals might not be able to sustain long-term metastatic latency. Tissues that host DTCs, such as the lungs, liver or bone marrow, do not exist in a perpetual state of growth inhibition. On the contrary, they support cell proliferation as part of normal tissue homeostasis and regeneration in a context that would regularly stimulate DTCs to enter the cell cycle. Therefore, cancer-cell-autonomous mechanisms that self-impose quiescence might be necessary to maintain DTCs in a dormant state. It is also unclear how a continuously quiescent DTC population could evolve and acquire the necessary traits for overt metastasis.

Evidence that DTCs are kept in a latent state by the immune system comes from the transmission of cancer during organ transplantation. Kidney, liver and lung transplants from donors who were cured of melanoma or who had developed glioblastoma, which is generally considered to be a nonmetastatic tumour, have resulted in the transmission of donor-derived tumours to immunosuppressed recipients^{110,111}. Such cases suggest that DTCs are maintained in a latent state by constant pressure from the immune system. DTCs might enter the cell cycle intermittently, and their progeny might undergo rapid elimination by the immune system while evolving to acquire traits that facilitate metastatic outbreak.

Overt metastasis

In some organs, breaking down growth-inhibitory or immune barriers can be sufficient to initiate the aggressive metastatic outgrowth of DTCs.

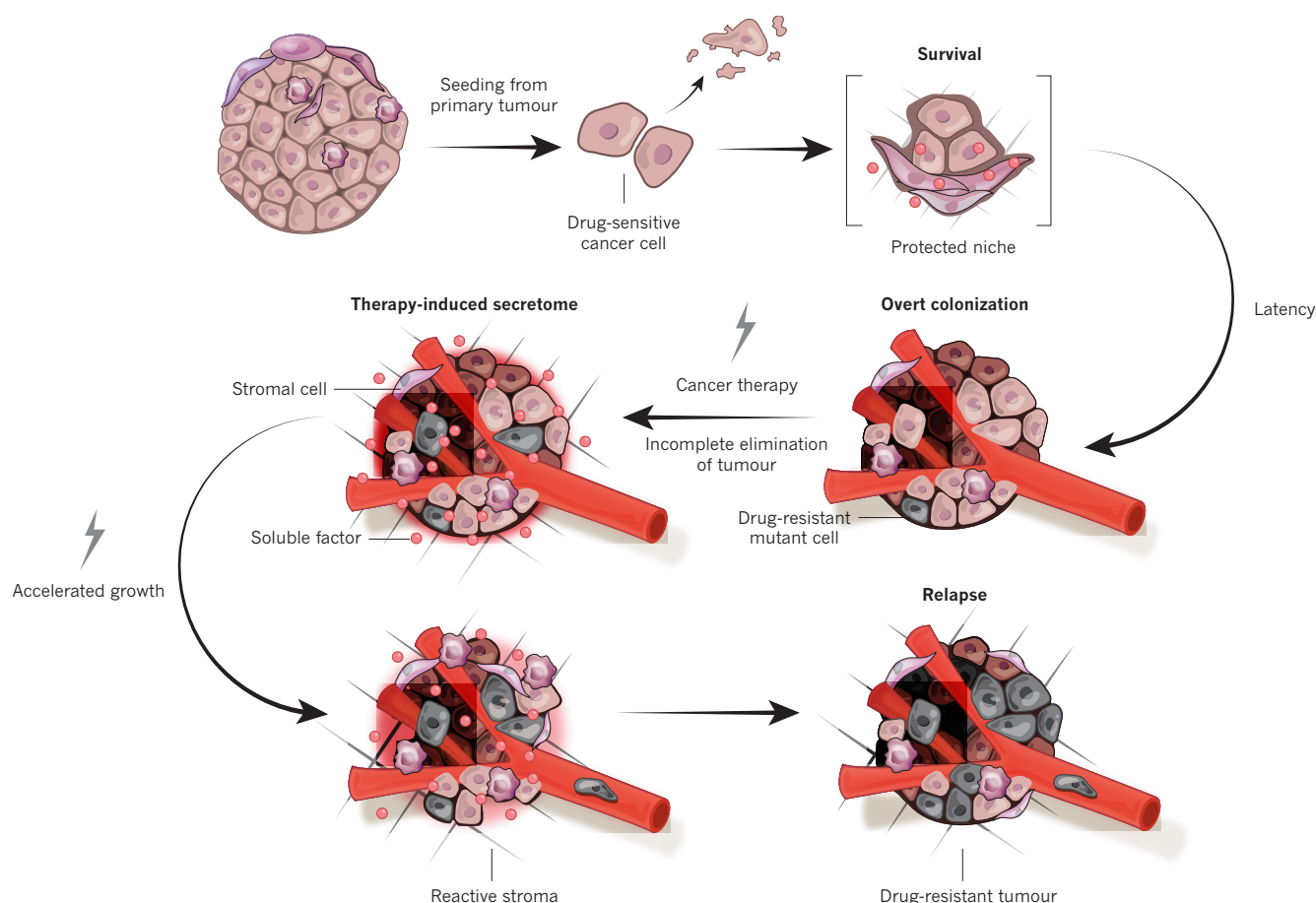


Figure 4 | The biology of metastasis, before and after cancer therapy.

Latent metastasis occurs owing to conditions that preserve the survival and tumour-initiating ability of disseminated cancer cells. The elimination of latent metastasis by targeting these survival mechanisms would prevent metastasis. Cancer cells that exit latency form manifest metastases. This condition is treated with combinations of conventional chemotherapy, targeted therapy and immunotherapy. Although treatment can dramatically reduce the metastatic burden, tumour elimination is frequently incomplete. A considerable

proportion of the tumour cell population will withstand treatment by adapting its intracellular pathways or by activating supportive paracrine inputs. The stress of targeted therapy causes drug-sensitive cancer cells to express a large number of secreted factors, known as a therapy-induced secretome, that can salvage drug-sensitive cells and accelerate the growth of minority drug-resistant cancer cells. This accelerated growth drives relapse as a drug-resistant tumour. The survival and growth mechanisms that residual cancer cells use during cancer therapy might resemble those used by their predecessors in the latent phase.

However, organs differ markedly in the structure and composition of their tissues, and their overt colonization involves distinct organ-specific metastatic traits¹¹². This translates into patterns of metastasis distribution that vary greatly depending on the tumour type. For example, prostate cancer has a propensity to relapse in bone; uveal melanoma tends to recur in the liver; and sarcomas often return in the lungs. In contrast, melanomas, breast carcinomas and lung adenocarcinomas tend to relapse in multiple organs. The kinetics of relapse also varies. For example, whereas recurrence in the brain frequently occurs at an early stage in metastatic lung cancer, this is typically a late event in metastatic breast cancer. Certain oncogenic mutations seem to affect metastatic tropism. For example, *KRAS*-mutant colon cancer advances to colonize the lungs from established liver metastases¹¹³.

Bone metastasis is the best-understood case of overt colonization and offers clear examples of the organ-specific traits that determine this final stage of metastasis. Osteolytic bone metastasis occurs when the balance of bone-generating and bone-resorbing osteoclasts is altered to favour the latter. Numerous mediators of osteoclast activation have been implicated in this process¹¹⁴. Cancer-cell-derived parathyroid hormone-related protein (PTHrP), IL-11 and TNF- α prompt osteoblasts to release receptor activator of NF κ B ligand (RANKL), a protein that stimulates osteoclast maturation^{114–116}. Bone metastatic cells also produce MMPs, which increase RANKL activity¹¹⁷ and reduce the levels of the RANKL antagonist osteoprotegerin¹¹⁸. Expression of the Notch ligand

Jagged1 and cell-adhesion molecules VCAM1 and soluble intercellular adhesion molecule 1 (sICAM1) also contributes to the mobilization of osteoclasts^{95,119,120}. Bone-matrix degradation by hyperactivated osteoclasts releases TGF- β that, in turn, augments the production of PTHrP, IL-11 and Jagged1 in the cancer cells and drives a vicious cycle of bone destruction^{114–116}.

Interestingly, prostate cancer cells that spread to bone alter the homeostatic balance in favour of osteoblastic activity, which stimulates the deposition of bone matrix and leads to eventual displacement of the bone marrow. Cancer-cell factors that are implicated in osteoblastic metastasis include fibroblast growth factors (FGFs), insulin-like growth factors (IGFs) and VEGFs, as well as endothelin 1, Wnt factors and BMPs¹¹⁴. Thus, bone metastasis provides a compelling example of how cancer cells engage their host microenvironment in overt metastasis. Similarly, specific stromal components can be engaged in other organs by metastatic cells with the necessary organ-specific colonization traits. For example, when breast and lung carcinomas spread to the brain, the cancer cells can profitably engage astrocytes and microglia by expressing endothelin 1 (ref. 121). However, our knowledge of overt-colonization traits that are specific to organs other than bone is limited, and these traits need further investigation.

In certain groups of patients, metastasis is confined mainly to a particular organ that is better able to resist therapy than others. An example is the rise in the incidence of late brain and leptomeningeal

metastasis in people with HER2⁺ breast cancer. Such patients can benefit from advances in targeted therapies that suppress extracranial metastasis. However, this success is often short-lived owing to the emergence of brain metastasis. Brain metastasis is a major cause of morbidity and mortality, with an overall incidence that is ten times higher than that of all primary brain tumours combined, and has few therapeutic options. A better understanding of its underlying mechanisms is urgently needed.

After therapy

The surgical removal of a malignant tumour is often complemented with radiotherapy and systemic chemotherapy to suppress relapse. If metastasis becomes clinically manifest, most systemic treatments are designed to target metastasis irrespective of organ site. Treatments include classic chemotherapy, targeted therapy against oncogenic drivers, immunotherapeutic agents that leverage the antitumour power of the immune system and, increasingly, a combination of all of these. Treatments that target metastasis in a particular organ — by taking aim at cancer-cell interactions with the host tissue — would be indicated when metastasis is confined to that organ, as is the case with bone metastasis in some people with breast cancer. A meta-analysis suggests that adjuvant therapy with osteoclast-inhibitory bisphosphonates suppresses bone metastasis and prolongs survival in postmenopausal women with breast cancer¹²². Denosumab, an antibody that targets RANKL, reduces the incidence of bone fractures associated with metastasis in patients who are receiving aromatase inhibitors¹²³.

Despite these advances, therapies frequently achieve only partial tumour shrinkage and leave behind substantial amounts of disease. Continued treatment can keep the residual tumour indolent for some time. However, drug-resistant cancer-cell clones eventually emerge that drive rapid relapse^{124,125}. As a result, the cure rates of patients with metastasis remain disappointingly low.

Research is beginning to focus on the biology of residual metastatic cells after therapy, and is aiming to better suppress the re-emergence of cancer cells (Fig. 4). The cancer-cell population can resist treatment through alterations to negative-feedback signalling loops¹²⁶ and supportive interactions with the tumour microenvironment. For example, DNA-damaging agents induce the secretion of trophic factors such as IL-6 and the metalloproteinase inhibitor Timp-1 in normal cells of the thymus, which creates a chemoprotective niche for the survival of residual cancer cells that facilitates eventual relapse¹²⁷. Similarly, stromal fibroblasts secrete Wnt16b in response to chemotherapy, which promotes resistance to therapy in prostate cancer¹²⁸. Chemotherapy-induced expression of TNF- α in tumour-associated endothelial and mesenchymal cells has been shown to amplify the expression of the pro-metastatic cytokine CXCL1 in cancer cells⁹². In BRAF-mutant melanomas that are treated with RAF inhibitors, tumour-associated macrophages secrete TNF- α and VEGF^{129,130}. Tumour-associated fibroblasts secrete hepatocyte growth factors (HGFs)¹³¹, which protect melanoma cells and limit the effectiveness of therapy.

Under the stress of therapy, the cancer cells themselves can be a source of survival signals^{54,132,133}. Targeted therapy with tyrosine-kinase inhibitors that is directed against melanoma (vemurafenib and dabrafenib) or lung adenocarcinoma (erlotinib and crizotinib) triggers the production of a complex secretome — the therapy-induced secretome — that activates multiple survival pathways in the remaining, drug-sensitive cancer cells⁵⁴. Furthermore, this secretome can stimulate the outgrowth, dissemination and further metastatic seeding of clones with mutations that confer drug resistance. Collectively, these findings reveal a complex biology in cancer cell populations that remain after treatment of metastatic tumours and contribute to tumour relapse.

Future directions

An important objective for current research is the identification of mediators of metastasis that are common to different organs and types of tumours. Although organ-specific metastasis has intrigued researchers for over a century, in reality many patients are affected by, or are at

risk of, metastasis in multiple organs. For these cases, it would be valuable to identify common mediators as potential therapeutic targets. For example, checkpoint immunotherapy — which has shown encouraging success in the clinic — is based on the premise that immune evasion is a shared feature of metastatic disease, irrespective of the organ site. Deeper understanding of the common mediators of metastasis and how tumours regrow after therapy would help to improve strategies for the elimination of residual disease.

The advent of single-cell analysis techniques, such as single-cell RNA sequencing and signalling-pathway profiling, is allowing functional and phenotypic analysis of heterogeneous cell populations in unprecedented detail^{134–137}. The application of these techniques to residual disease and overt metastases will enable researchers to define tumour heterogeneity, cell-population structures and evolution, and cell-type-specific response patterns to stromal cues and therapeutic agents, as well as other parameters, at a depth never before possible. Furthermore, the ability to analyse circulating tumour DNA in the blood of people with cancer will facilitate the monitoring of therapy responses, the emergence of distinct resistant clones and patterns of early disease recurrence.

Preventing metastasis in high-risk patients would be far better than having to treat it. The systemic nature of metastatic disease, the heterogeneity of metastatic tumours, the multitude of genes and pathways involved in different organs and the many mechanisms of drug resistance paint a sobering picture of the problems that must be overcome to address overt metastatic disease. Ostensibly, the goal of systemic therapy that is delivered after removal of a primary tumour is to prevent relapse. However, most agents used in adjuvant therapy are designed to target growing cancer cells rather than the quiescent DTCs that predominate during metastatic latency. A better understanding of the basis for metastatic colonization and its latent phase, in particular, is needed to develop superior treatments. Research on the mechanisms that support the viability of latent metastatic cells should yield clues for targeting residual disease, with the goal of preventing metastasis. ■

Received 26 August; accepted 11 November 2015.

1. Nagrath, S. *et al.* Isolation of rare circulating tumour cells in cancer patients by microchip technology. *Nature* **450**, 1235–1239 (2007).
 2. Braun, S. *et al.* A pooled analysis of bone marrow micrometastasis in breast cancer. *N. Engl. J. Med.* **353**, 793–802 (2005).
 3. Wong, C. W. *et al.* Apoptosis: an early event in metastatic inefficiency. *Cancer Res.* **61**, 333–338 (2001).
 4. Minn, A. J. *et al.* Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors. *J. Clin. Invest.* **115**, 44–55 (2005).
 5. Chambers, A. F., Groom, A. C. & MacDonald, I. C. Dissemination and growth of cancer cells in metastatic sites. *Nature Rev. Cancer* **2**, 563–572 (2002).
 6. Luzzi, K. J. *et al.* Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *Am. J. Pathol.* **153**, 865–873 (1998).
 7. Cameron, M. D. *et al.* Temporal progression of metastasis in lung: cell survival, dormancy, and location dependence of metastatic inefficiency. *Cancer Res.* **60**, 2541–2546 (2000).
 8. Kienast, Y. *et al.* Real-time imaging reveals the single steps of brain metastasis formation. *Nature Med.* **16**, 116–122 (2010).
 9. Valiente, M. *et al.* Serpins promote cancer cell survival and vascular co-option in brain metastasis. *Cell* **156**, 1002–1016 (2014).
 10. Heyn, C. *et al.* *In vivo* MRI of cancer cell fate at the single-cell level in a mouse model of breast cancer metastasis to the brain. *Magn. Reson. Med.* **56**, 1001–1010 (2006).
 11. Malanchi, I. *et al.* Interactions between cancer stem cells and their niche govern metastatic colonization. *Nature* **481**, 85–89 (2012).
 12. Calon, A. *et al.* Dependency of colorectal cancer on a TGF- β -driven program in stromal cells for metastasis initiation. *Cancer Cell* **22**, 571–584 (2012).
- This study used multiphoton laser-scanning microscopy to track the fate of individual metastasizing cancer cells in the brain.**
- This study identified that proteins called serpins promote metastasis in the brain by shielding cancer cells from the tissue defences of the reactive brain stroma and also promote vascular co-option mediated by LICAM.**
- This study demonstrated that breast cancer stem cells generate a permissive metastatic niche by triggering the secretion of stromal periostin.**
- This study found that TGF- β signalling in stromal cells increases the efficiency of organ colonization by colorectal cancer cells and shed light on the paradox that high levels of TGF- β are associated with poorer prognosis in colorectal cancer, despite the fact that cancer cells frequently exhibit inactivating mutations in TGF- β -pathway components.**

13. Nguyen, D. X., Bos, P. D. & Massague, J. Metastasis: from dissemination to organ-specific colonization. *Nature Rev. Cancer* **9**, 274–284 (2009).
14. Chaffer, C. L. & Weinberg, R. A. A perspective on cancer cell metastasis. *Science* **331**, 1559–1564 (2011).
15. Hall, A. The cytoskeleton and cancer. *Cancer Metastasis Rev.* **28**, 5–14 (2009).
16. Kessenbrock, K., Plaks, V. & Werb, Z. Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell* **141**, 52–67 (2010).
17. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nature Med.* **19**, 1423–1437 (2013).
18. Giampieri, S. *et al.* Localized and reversible TGF β signalling switches breast cancer cells from cohesive to single cell motility. *Nature Cell Biol.* **11**, 1287–1296 (2009).
19. Roh-Johnson, M. *et al.* Macrophage contact induces RhoA GTPase signaling to trigger tumor cell intravasation. *Oncogene* **33**, 4203–4212 (2014).
20. Friedl, P. & Gilmour, D. Collective cell migration in morphogenesis, regeneration and cancer. *Nature Rev. Mol. Cell Biol.* **10**, 445–457 (2009).
21. Magnon, C. *et al.* Autonomic nerve development contributes to prostate cancer progression. *Science* **341**, 1236361 (2013).
22. Thiery, J. P., Acloque, H., Huang, R. Y. & Nieto, M. A. Epithelial-mesenchymal transitions in development and disease. *Cell* **139**, 871–890 (2009).
23. Tam, W. L. & Weinberg, R. A. The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nature Med.* **19**, 1438–1449 (2013).
24. Fischer, K. R. *et al.* Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**, 472–476 (2015).
25. Zheng, X. *et al.* Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **527**, 525–530 (2015).
26. Tabassum, D. P. & Polyak, K. Tumorigenesis: it takes a village. *Nature Rev. Cancer* **15**, 473–483 (2015).
27. Marusyk, A. *et al.* Non-cell-autonomous driving of tumour growth supports subclonal heterogeneity. *Nature* **514**, 54–58 (2014).
This study showed that a minority subclone in a heterogeneous tumour is able to drive proliferation of the whole tumour by overcoming environmental constraints.
28. Cleary, A. S., Leonard, T. L., Gestl, S. A. & Gunther, E. J. Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers. *Nature* **508**, 113–117 (2014).
29. Calbo, J. *et al.* A functional role for tumor cell heterogeneity in a mouse model of small cell lung cancer. *Cancer Cell* **19**, 244–256 (2011).
30. Gundem, G. *et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* **520**, 353–357 (2015).
This study used whole-genome sequencing and computational analysis to provide evidence for metastasis-to-metastasis spread and the transfer of multiple tumour clones between metastatic sites in people with prostate cancer.
31. Aceto, N. *et al.* Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* **158**, 1110–1122 (2014).
This study showed that CTC clusters have increased metastatic potential compared with single CTCs.
32. Labelle, M., Begum, S. & Hynes, R. O. Direct signaling between platelets and cancer cells induces an epithelial-mesenchymal-like transition and promotes metastasis. *Cancer Cell* **20**, 576–590 (2011).
33. Le Gal, K. *et al.* Antioxidants can increase melanoma metastasis in mice. *Sci. Transl. Med.* **7**, 308re8 (2015).
34. Piskounova, E. *et al.* Oxidative stress inhibits distant metastasis by human melanoma cells. *Nature* **527**, 186–191 (2015).
35. Denève, E. *et al.* Capture of viable circulating tumor cells in the liver of colorectal cancer patients. *Clin. Chem.* **59**, 1384–1392 (2013).
36. Al-Mehdi, A. B. *et al.* Intravascular origin of metastasis from the proliferation of endothelium-attached tumor cells: a new model for metastasis. *Nature Med.* **6**, 100–102 (2000).
37. Aird, W. C. Phenotypic heterogeneity of the endothelium: I. Structure, function, and mechanisms. *Circ. Res.* **100**, 158–173 (2007).
38. Budczies, J. *et al.* The landscape of metastatic progression patterns across major human cancers. *Oncotarget* **6**, 570–583 (2015).
39. Minn, A. J. *et al.* Genes that mediate breast cancer metastasis to lung. *Nature* **436**, 518–524 (2005).
40. Matsuda, Y., Schlange, T., Oakeley, E. J., Boulay, A. & Hynes, N. E. WNT signaling enhances breast cancer cell motility and blockade of the WNT pathway by sFRP1 suppresses MDA-MB-231 xenograft growth. *Breast Cancer Res.* **11**, R32 (2009).
41. Padua, D. *et al.* TGF β primes breast tumors for lung metastasis seeding through angiopoietin-like 4. *Cell* **133**, 66–77 (2008).
42. Gupta, G. P. *et al.* Mediators of vascular remodelling co-opted for sequential steps in lung metastasis. *Nature* **446**, 765–770 (2007).
43. Tichet, M. *et al.* Tumour-derived SPARC drives vascular permeability and extravasation through endothelial VCAM1 signalling to promote metastasis. *Nature Commun.* **6**, 6993 (2015).
44. Weis, S., Cui, J., Barnes, L. & Cheresh, D. Endothelial barrier disruption by VEGF-mediated Src activity potentiates tumor cell extravasation and metastasis. *J. Cell Biol.* **167**, 223–229 (2004).
45. Schumacher, D., Striic, B., Sivaraj, K. K., Wettschreck, N. & Offermanns, S. Platelet-derived nucleotides promote tumor-cell transendothelial migration and metastasis via P2Y2 receptor. *Cancer Cell* **24**, 130–137 (2013).
46. Qian, B. Z. *et al.* CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. *Nature* **475**, 222–225 (2011).
47. Bos, P. D. *et al.* Genes that mediate breast cancer metastasis to the brain. *Nature* **459**, 1005–1009 (2009).
48. Sevenich, L. *et al.* Analysis of tumour- and stroma-supplied proteolytic networks reveals a brain-metastasis-promoting role for cathepsin S. *Nature Cell Biol.* **16**, 876–888 (2014).
49. Zhou, W. *et al.* Cancer-secreted miR-105 destroys vascular endothelial barriers to promote metastasis. *Cancer Cell* **25**, 501–515 (2014).
50. Tominaga, N. *et al.* Brain metastatic cancer cells release microRNA-181c-containing extracellular vesicles capable of destructing blood-brain barrier. *Nature Commun.* **6**, 6716 (2015).
51. Hanahan, D. & Coussens, L. M. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell* **21**, 309–322 (2012).
52. Fridly, I. J. The pathogenesis of cancer metastasis: the ‘seed and soil’ hypothesis revisited. *Nature Rev. Cancer* **3**, 453–458 (2003).
53. Kim, M. Y. *et al.* Tumor self-seeding by circulating cancer cells. *Cell* **139**, 1315–1326 (2009).
54. Obenaus, A. C. *et al.* Therapy-induced tumour secretomes promote resistance and tumour progression. *Nature* **520**, 368–372 (2015).
55. Eyles, J. *et al.* Tumor cells disseminate early, but immunosurveillance limits metastatic outgrowth, in a mouse model of melanoma. *J. Clin. Invest.* **120**, 2030–2039 (2010).
56. Bidwell, B. N. *et al.* Silencing of Irf7 pathways in breast cancer cells promotes bone metastasis through immune escape. *Nature Med.* **18**, 1224–1231 (2012).
57. Smyth, M. J. *et al.* Perforin is a major contributor to NK cell control of tumor metastasis. *J. Immunol.* **162**, 6658–6662 (1999).
58. Paolino, M. *et al.* The E3 ligase Cbl-b and TAM receptors regulate cancer metastasis via natural killer cells. *Nature* **507**, 508–512 (2014).
59. Milsom, C. C., Lee, C. R., Hackl, C., Man, S. & Kerbel, R. S. Differential post-surgical metastasis and survival in SCID, NOD-SCID and NOD-SCID-IL-2Ry^{null} mice with parental and syngeneic variants of human breast cancer: implications for host defense mechanisms regulating metastasis. *PLoS ONE* **8**, e71270 (2013).
60. Takeda, K. *et al.* Involvement of tumor necrosis factor-related apoptosis-inducing ligand in surveillance of tumor metastasis by liver natural killer cells. *Nature Med.* **7**, 94–100 (2001).
61. Postow, M. A. *et al.* Nivolumab and ipilimumab versus ipilimumab in untreated melanoma. *N. Engl. J. Med.* **372**, 2006–2017 (2015).
62. Sharma, P. & Allison, J. P. Immune checkpoint targeting in cancer therapy: toward combination strategies with curative potential. *Cell* **161**, 205–214 (2015).
63. Morrison, S. J. & Spradling, A. C. Stem cells and niches: mechanisms that promote stem cell maintenance throughout life. *Cell* **132**, 598–611 (2008).
64. Hsu, Y. C., Li, L. & Fuchs, E. Emerging interactions between skin stem cells and their niches. *Nature Med.* **20**, 847–856 (2014).
65. Schepers, A. G. *et al.* Lineage tracing reveals Lgr5⁺ stem cell activity in mouse intestinal adenomas. *Science* **337**, 730–735 (2012).
66. Chen, J. *et al.* A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature* **488**, 522–526 (2012).
67. Driessens, G., Beck, B., Caauwe, A., Simons, B. D. & Blanpain, C. Defining the mode of tumour growth by clonal analysis. *Nature* **488**, 527–530 (2012).
68. Kreso, A. & Dick, J. E. Evolution of the cancer stem cell model. *Cell Stem Cell* **14**, 275–291 (2014).
69. Oskarsson, T., Batlle, E. & Massague, J. Metastatic stem cells: sources, niches, and vital pathways. *Cell Stem Cell* **14**, 306–321 (2014).
70. Shiozawa, Y. *et al.* Human prostate cancer metastases target the hematopoietic stem cell niche to establish footholds in mouse bone marrow. *J. Clin. Invest.* **121**, 1298–1312 (2011).
71. Müller, A. *et al.* Involvement of chemokine receptors in breast cancer metastasis. *Nature* **410**, 50–56 (2001).
72. Zhang, X. H. *et al.* Selection of bone metastasis seeds by mesenchymal signals in the primary tumor stroma. *Cell* **154**, 1060–1073 (2013).
This study showed how stromal signals that resemble those in the bone marrow can skew heterogeneous cancer cell populations in the primary tumour towards a predominance of clones that are primed for bone metastasis.
73. Hambardzumyan, D. *et al.* PI3K pathway regulates survival of cancer stem cells residing in the perivascular niche following radiation in medulloblastoma *in vivo*. *Genes Dev.* **22**, 436–448 (2008).
74. Cao, Z. *et al.* Angiocrine factors deployed by tumor vascular niche induce B cell lymphoma invasiveness and chemoresistance. *Cancer Cell* **25**, 350–365 (2014).
75. Carbonell, W. S., Anson, O., Sibson, N. & Muschel, R. The vascular basement membrane as “soil” in brain metastasis. *PLoS ONE* **4**, e5857 (2009).
76. Oskarsson, T. *et al.* Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs. *Nature Med.* **17**, 867–874 (2011).
77. Erler, J. T. *et al.* Lysyl oxidase is essential for hypoxia-induced metastasis. *Nature* **440**, 1222–1226 (2006).
78. Gilkes, D. M. *et al.* Procollagen lysyl hydroxylase 2 is essential for hypoxia-induced breast cancer metastasis. *Mol. Cancer Res.* **11**, 456–466 (2013).
79. Eisinger-Mathason, T. S. *et al.* Hypoxia-dependent modification of collagen networks promotes sarcoma metastasis. *Cancer Discov.* **3**, 1190–1205 (2013).
80. Kaplan, R. N. *et al.* VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche. *Nature* **438**, 820–827 (2005).
81. McAllister, S. S. & Weinberg, R. A. The tumour-induced systemic environment as a critical regulator of cancer progression and metastasis. *Nature Cell Biol.* **16**, 717–727 (2014).
82. Cox, T. R. *et al.* The hypoxic cancer secretome induces pre-metastatic bone lesions through lysyl oxidase. *Nature* **522**, 106–110 (2015).

83. Costa-Silva, B. *et al.* Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nature Cell Biol.* **17**, 816–826 (2015).
84. Erler, J. T. *et al.* Hypoxia-induced lysyl oxidase is a critical mediator of bone marrow cell recruitment to form the premetastatic niche. *Cancer Cell* **15**, 35–44 (2009).
85. Wculek, S. K. & Malanchi, I. Neutrophils support lung colonization of metastasis-initiating breast cancer cells. *Nature* <http://dx.doi.org/10.1038/nature16140> (9 December 2015).
86. Peinado, H. *et al.* Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. *Nature Med.* **18**, 883–891 (2012).
87. Hoshino, A. *et al.* Tumour exosome integrins determine organotropic metastasis. *Nature* **527**, 329–335 (2015).
88. Nowak, D. G. *et al.* MYC drives Pten/Trp53-deficient proliferation and metastasis due to IL6 secretion and AKT suppression via PHLPP2. *Cancer Discov.* **5**, 636–651 (2015).
89. Li, B. *et al.* Id1-induced IGF-II and its autocrine/endocrine promotion of esophageal cancer progression and chemoresistance—implications for IGF-II and IGF-IR-targeted therapy. *Clin. Cancer Res.* **20**, 2651–2662 (2014).
90. Zhang, X. H. *et al.* Latent bone metastasis in breast cancer tied to Src-dependent survival signals. *Cancer Cell* **16**, 67–78 (2009).
91. Wyckoff, J. *et al.* A paracrine loop between tumor cells and macrophages is required for tumor cell migration in mammary tumors. *Cancer Res.* **64**, 7022–7029 (2004).
92. Acharyya, S. *et al.* A CXCL1 paracrine network links cancer chemoresistance and metastasis. *Cell* **150**, 165–178 (2012).
93. Tabariès, S. *et al.* Claudin-2 promotes breast cancer liver metastasis by facilitating tumor cell interactions with hepatocytes. *Mol. Cell. Biol.* **32**, 2979–2991 (2012).
94. Chen, Q., Zhang, X. H. & Massague, J. Macrophage binding to receptor VCAM-1 transmits survival signals in breast cancer cells that invade the lungs. *Cancer Cell* **20**, 538–549 (2011).
95. Lu, X. *et al.* VCAM-1 promotes osteolytic expansion of indolent bone micrometastasis of breast cancer by engaging $\alpha 4 \beta 1$ -positive osteoclast progenitors. *Cancer Cell* **20**, 701–714 (2011).
96. Vanharanta, S. *et al.* Epigenetic expansion of VHL-HIF signal output drives multiorgan metastasis in renal cancer. *Nature Med.* **19**, 50–56 (2013).
97. Tavazoie, S. F. *et al.* Endogenous human microRNAs that suppress breast cancer metastasis. *Nature* **451**, 147–152 (2008).
98. Ma, L., Teruya-Feldstein, J. & Weinberg, R. A. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature* **449**, 682–688 (2007).
99. Korpai, M. *et al.* Direct targeting of Sec23a by miR-200s influences cancer cell secretome and promotes metastatic colonization. *Nature Med.* **17**, 1101–1108 (2011).
100. Pencheva, N. *et al.* Convergent multi-miRNA targeting of ApoE drives LRP1/LRP8-dependent melanoma metastasis and angiogenesis. *Cell* **151**, 1068–1082 (2012).
101. Pantel, K. & Brakenhoff, R. H. Dissecting the metastatic cascade. *Nature Rev. Cancer* **4**, 448–456 (2004).
102. Sosa, M. S., Bragado, P. & Aguirre-Ghiso, J. A. Mechanisms of disseminated cancer cell dormancy: an awakening field. *Nature Rev. Cancer* **14**, 611–622 (2014).
103. Kang, Y. & Pantel, K. Tumor cell dissemination: emerging biological insights from animal models and cancer patients. *Cancer Cell* **23**, 573–581 (2013).
104. Gao, H. *et al.* The BMP inhibitor Coco reactivates breast cancer cells at lung metastatic sites. *Cell* **150**, 764–779 (2012).
105. Bragado, P. *et al.* TGF- $\beta 2$ dictates disseminated tumour cell fate in target organs through TGF- β -RIII and p38 α / β signalling. *Nature Cell Biol.* **15**, 1351–1361 (2013).
106. Kobayashi, A. *et al.* Bone morphogenetic protein 7 in dormancy and metastasis of prostate cancer stem-like cells in bone. *J. Exp. Med.* **208**, 2641–2655 (2011).
107. Ghajar, C. M. *et al.* The perivascular niche regulates breast tumour dormancy. *Nature Cell Biol.* **15**, 807–817 (2013).
108. Barkan, D. *et al.* Metastatic growth from dormant cells induced by a col-I-enriched fibrotic environment. *Cancer Res.* **70**, 5706–5716 (2010).
109. Aguirre-Ghiso, J. A., Liu, D., Mignatti, A., Kovalski, K. & Ossowski, L. Urokinase receptor and fibronectin regulate the ERKMAPK to p38MAPK activity ratios that determine carcinoma cell proliferation or dormancy *in vivo*. *Mol. Biol. Cell* **12**, 863–879 (2001).
110. Strauss, D. C. & Thomas, J. M. Transmission of donor melanoma by organ transplantation. *Lancet Oncol.* **11**, 790–796 (2010).
111. Collignon, F. P., Holland, E. C. & Feng, S. Organ donors with malignant gliomas: an update. *Am. J. Transplant.* **4**, 15–21 (2004).
112. Obenauf, A. C. & Massagué, J. Surviving at a distance: organ-specific metastasis. *Trends Cancer* **1**, 76–91 (2015).
113. Urošević, J. *et al.* Colon cancer cells colonize the lung from established liver metastases through p38 MAPK signalling and PTHLH. *Nature Cell Biol.* **16**, 685–694 (2014).
114. Weilbaecher, K. N., Guise, T. A. & McCauley, L. K. Cancer to bone: a fatal attraction. *Nature Rev. Cancer* **11**, 411–425 (2011).
115. Yin, J. J. *et al.* TGF- β signaling blockade inhibits PTHrP secretion by breast cancer cells and bone metastases development. *J. Clin. Invest.* **103**, 197–206 (1999).
116. Kang, Y. *et al.* A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* **3**, 537–549 (2003).
117. Lynch, C. C. *et al.* MMP-7 promotes prostate cancer-induced osteolysis via the solubilization of RANKL. *Cancer Cell* **7**, 485–496 (2005).
118. Lu, X. *et al.* ADAMTS1 and MMP1 proteolytically engage EGF-like ligands in an osteolytic signaling cascade for bone metastasis. *Genes Dev.* **23**, 1882–1894 (2009).
119. Sethi, N., Dai, X., Winter, C. G. & Kang, Y. Tumor-derived JAGGED1 promotes osteolytic bone metastasis of breast cancer by engaging notch signaling in bone cells. *Cancer Cell* **19**, 192–205 (2011).
120. Ell, B. *et al.* Tumor-induced osteoclast miRNA changes as regulators and biomarkers of osteolytic bone metastasis. *Cancer Cell* **24**, 542–556 (2013).
121. Kim, S. W. *et al.* Role of the endothelin axis in astrocyte- and endothelial cell-mediated chemoprotection of cancer cells. *Neuro Oncol.* **16**, 1585–1598 (2014).
122. Early Breast Cancer Trialists' Collaborative Group (EBCTCG) *et al.* Adjuvant bisphosphonate treatment in early breast cancer: meta-analyses of individual patient data from randomised trials. *Lancet* **386**, 1353–1361 (2015).
123. Gnant, M. *et al.* Adjuvant denosumab in breast cancer (ABCSG-18): a multicentre, randomised, double-blind, placebo-controlled trial. *Lancet* **386**, 433–443 (2015).
124. Holohan, C., Van Schaeybroeck, S., Longley, D. B. & Johnston, P. G. Cancer drug resistance: an evolving paradigm. *Nature Rev. Cancer* **13**, 714–726 (2013).
125. Higgins, M. J. & Baselga, J. Targeted therapies for breast cancer. *J. Clin. Invest.* **121**, 3797–3803 (2011).
126. Lito, P. *et al.* Relief of profound feedback inhibition of mitogenic signaling by RAF inhibitors attenuates their activity in BRAFV600E melanomas. *Cancer Cell* **22**, 668–682 (2012).
127. Gilbert, L. A. & Hemann, M. T. DNA damage-mediated induction of a chemoresistant niche. *Cell* **143**, 355–366 (2010).
- This study shows that in response to DNA damage, paracrine factors are released in the thymus, which creates a chemoresistant niche that promotes the survival of a minimal residual tumour burden and serves as a reservoir for eventual relapse in a mouse model of Burkitt's lymphoma.**
128. Sun, Y. *et al.* Treatment-induced damage to the tumor microenvironment promotes prostate cancer therapy resistance through WNT16B. *Nature Med.* **18**, 1359–1368 (2012).
129. Smith, M. P. *et al.* The immune microenvironment confers resistance to MAPK pathway inhibitors through macrophage-derived TNF α . *Cancer Discov.* **4**, 1214–1229 (2014).
130. Wang, T. *et al.* BRAF inhibition stimulates melanoma-associated macrophages to drive tumor growth. *Clin. Cancer Res.* **21**, 1652–1664 (2015).
131. Straussman, R. *et al.* Tumour micro-environment elicits innate resistance to RAF inhibitors through HGF secretion. *Nature* **487**, 500–504 (2012).
132. Kurtova, A. V. *et al.* Blocking PGE2-induced tumour repopulation abrogates bladder cancer chemoresistance. *Nature* **517**, 209–213 (2015).
133. Huang, Q. *et al.* Caspase 3-mediated stimulation of tumor cell repopulation during cancer radiotherapy. *Nature Med.* **17**, 860–866 (2011).
134. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
135. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
136. Levine, J. H. *et al.* Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* **162**, 184–197 (2015).
137. Lawson, D. A. *et al.* Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature* **526**, 131–135 (2015).
- Using single-cell gene-expression profiling, this study found that early metastatic cancer cells possess a stem-like gene expression signature and give rise to heterogeneous tumours, which provides evidence to support a hierarchical metastasis stem-cell model.**
138. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
139. Vanharanta, S. & Massagué, J. Origins of metastatic traits. *Cancer Cell* **24**, 410–421 (2013).
140. Naxerova, K. & Jain, R. K. Using tumour phylogenetics to identify the roots of metastasis in humans. *Nature Rev. Clin. Oncol.* **12**, 258–272 (2015).
141. Campbell, P. J. *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**, 1109–1113 (2010).
142. Robinson, D. *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228 (2015).
143. Pencheva, N. & Tavazoie, S. F. Control of metastatic progression by microRNA regulatory networks. *Nature Cell Biol.* **15**, 546–554 (2013).
144. Klein, C. A. Parallel progression of primary tumours and metastases. *Nature Rev. Cancer* **9**, 302–312 (2009).

Acknowledgements The authors thank K. Ganesh and T. Wiesner for useful input. J.M. was supported by US National Institutes of Health grants CA163167 and CA129243, the Congressionally Directed Medical Research Programs of the US Department of Defense, Cancer Center Support Grant P30 CA008748 and the Center for Metastasis Research of the Memorial Sloan Kettering Cancer Center. A.C.O. was an Erwin Schrödinger Fellowship awardee (J3013, Austrian Science Fund).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this paper at go.nature.com/fi9jfd. Correspondence should be addressed to J.M. (j-massague@ski.mskcc.org).

Reparative inflammation takes charge of tissue regeneration

Michael Karin¹ & Hans Clevers^{2,3}

Inflammation underlies many chronic and degenerative diseases, but it also mitigates infections, clears damaged cells and initiates tissue repair. Many of the mechanisms that link inflammation to damage repair and regeneration in mammals are conserved in lower organisms, indicating that it is an evolutionarily important process. Recent insights have shed light on the cellular and molecular processes through which conventional inflammatory cytokines and Wnt factors control mammalian tissue repair and regeneration. This is particularly important for regeneration in the gastrointestinal system, especially for intestine and liver tissues in which aberrant and deregulated repair results in severe pathologies.

The intestinal epithelium is the most rapidly self-renewing tissue in the mammalian body — cells have a life cycle of 3–4 days. Multiple studies have demonstrated that the intestinal stem cells (ISCs) in intestinal crypt compartments are dependent on Wnt homeostatic signals¹. ISCs are marked by the expression of the transmembrane receptor *Lgr5* and reside at the bottom of the crypts, where they are intermingled with Paneth cells, one of their daughter cells. Paneth cells produce bactericidal proteins (for example, lysozyme) and peptides (defensins and cryptdins), thus participating in gut innate immunity (Fig. 1a). Paneth cells also produce a range of niche signals that support ISCs, including epidermal growth factor (EGF), Wnt3 and Notch ligands². Although Wnt3 is redundant with other Wnts *in vivo*³, it is essential for the expansion of ISC-generated organoids *in vitro*¹.

The healthy liver contains very few proliferative cells. Increased Wnt activity mainly occurs around the liver's central veins, resulting in a signalling gradient that governs 'zonation' of the liver lobule — the differential distribution of liver enzymes and metabolic functions along the portal-tract-to-central-vein axis⁴. A pool of pericentral diploid hepatocytes has been described using lineage-tracing based on the Wnt-driven adult stem cell (ASC) gene *Axin2*. This cell population mediates homeostatic hepatocyte renewal and is controlled by Wnts produced by endothelial cells adjacent to the central vein⁵. However, the liver has two different and powerful regenerative responses when damaged⁶. Partial hepatectomy (when up to 65% of the organ is removed) activates a unique response, during which the remaining, fully differentiated, healthy diploid hepatocytes enter the cell cycle and, within a matter of weeks, restore the liver mass to its original size⁷. A number of the inflammatory signals described in this Review play a major part in this type of liver regeneration by acting on mature hepatocytes, but Wnt signals are also thought to have a role^{8,9}. A very different pattern of regeneration, often given the pathological description oval-cell response or ductular reaction^{10,11}, is elicited by insults that debilitate all hepatocytes, such as exposure to liver toxins, viral infections or immune attack. Oval cells are postulated to be bipotent stem cells (they have the capacity to generate hepatocytes and biliary cells) that are derived from the biliary tract system (the canals of Hering; Fig. 1b). The paucity of unique markers for oval cells and their apparent absence in the healthy liver has

complicated mechanistic studies of these elusive stem cells, and their relevance remains a subject of intense debate (see later).

Inflammation, and normal and abnormal damage repair

People often react to inflammation and its five signs: *dolor* (pain), *calor* (heat), *rubor* (redness), *tumour* (swelling) and *functio laesa* (loss of function) by taking an anti-inflammatory medication. But inflammation is an important protective response that, along with the elimination of its primary triggers (foreign organisms, dead cells or physical irritants), plays a crucial part in the regeneration of injured tissues. Too little inflammation can result in tissue destruction by harmful triggers, especially bacteria, whereas chronic unresolved inflammation culminates in a host of pathologies, including cancer and fibrosis. The link between inflammation and cancer is reviewed extensively elsewhere^{12,13}, as is the link between inflammation and fibrosis¹⁴. Both of these pathologies can be viewed as attempts at tissue repair that have gone awry. In this Review, we argue that self-limiting acute inflammation is essential for a proper restorative response, and we focus on this topic. Wound healing and injury repair facilitate the resolution of inflammation by restoring barrier function. Self-resolving inflammation is the first stage of wound repair and is followed by tissue formation and eventual remodelling¹⁵. Although not as extensively studied as its innate immune or antimicrobial functions, the regenerative function of inflammation is evolutionarily conserved and has been amply documented in the fruit fly *Drosophila*, in which genetic analysis has highlighted its role in regeneration of adult tissues such as the injured midgut¹⁶ and in closure of larval wounds¹⁷.

Evolutionarily conserved repair pathways in the fly gut

The fly innate immune system is activated on engagement of pattern recognition receptors (PRRs) by pathogen-associated molecular patterns (PAMPs), setting in motion highly conserved signalling cascades that impinge on NF- κ B, AP-1 and STAT transcription factors, which are also the main regulators of the mammalian inflammatory response¹⁸. In innate immunity, these pathways are important for the induction of antimicrobial peptides in haemocytes, the fat body and the midgut — the fly equivalents of myeloid cells, the liver and the mammalian intestine, respectively. The same pathways also control regeneration and wound healing by stimulating proliferation of ISCs and ASCs, and modulating

¹Laboratory of Gene Regulation and Signal Transduction, Departments of Pharmacology and Pathology, Moores Cancer Center, University of California San Diego School of Medicine, 9500 Gilman Drive, La Jolla, California 92093-0636, USA. ²Princess Máxima Center and Hubrecht Institute, Uppsalalaan 8, 3584 CR Utrecht, the Netherlands. ³University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, the Netherlands.

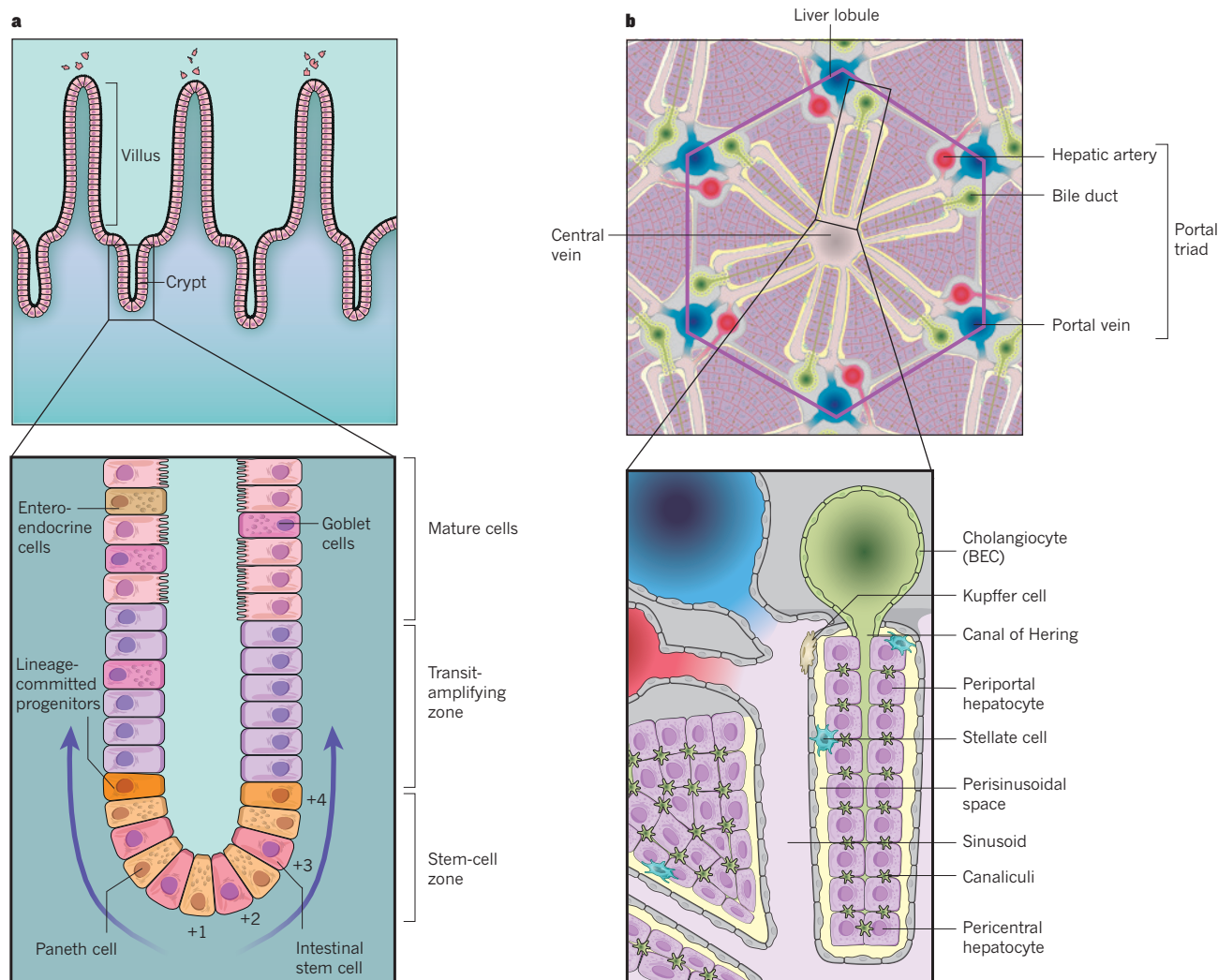


Figure 1 | The microanatomy of the cellular compartments responsible for intestinal and liver regeneration. **a**, Intestinal stem cells (ISCs) are located at the bottom of crypts in the stem-cell zone. They sit wedged between Paneth cells, which provide niche factors. The stem-cell positions are denoted by +1 to +3 from the bottom of the crypt. The label-retaining cells at the +4 position can act as reserve stem cells when crypt damage occurs. The transit-amplifying zone above the stem-cell zone contains rapidly proliferating lineage-committed progenitors that mature as they move up the crypt–villus axis. On injury, ISCs expand and repair the mucosa to restore the gut permeability barrier. **b**, The liver is made up of roughly hexagonal functional units called lobules. Within these, cords of hepatocytes stretch from the central vein to the portal triad. Hepatocytes differ in function, depending on their position (pericentral to periportal). On the apical side, hepatocytes enclose canaliculi, bile channels that lead

outwards to intrahepatic bile ducts. Along the way, bile passes through the canal of Hering, which is lined by hepatocytes and cholangiocytes. Basolaterally, hepatocytes face the perisinusoidal space. This transitional compartment between hepatocyte and sinusoid is delimited by discontinuous, fenestrated endothelial cells and contains hepatic stellate cells. Kupffer cells, a specialized population of liver macrophages, reside at the interface between the perisinusoidal space and sinusoid. Blood from the portal vein and the hepatic artery mixes in sinusoids and flows towards the central vein, transporting both oxygen from the lungs and nutrients from the digestive tract to the liver. Although liver stem cells have not been identified, pericentral diploid hepatocytes were suggested to be responsible for homeostatic hepatocyte renewal. Chronic liver injury, however, leads to expansion of periportal hybrid hepatocytes, which repair the injured liver parenchyma.

their differentiation¹⁶. Reactive oxygen species (ROS) produced during tissue injury and infection are an important cue that couples inflammation to ISC proliferation through activation of Jun N-terminal kinases (JNKs) and the antioxidant transcription factor and NRF2 homologue CncC¹⁹. JNK stimulates ISC proliferation by activating Fos (AP-1), which is also activated by growth factors of the EGF family. In addition, JNK contributes to the induction of Upd (a homologue of interleukin (IL)-6) family members that activate JAK–STAT signalling in both ISCs and visceral muscle cells, in which it also induces expression of growth factors (EGF family members) that directly stimulate ISC proliferation¹⁹. JNK activation also has a crucial role in epithelial sheet movement and cell migration, which are the first steps in wound closure. Completion of the regenerative and wound-healing response depends on compensatory proliferation of activated ASCs, normal differentiated

cells (NDCs) or dedifferentiated cells that assume a cellular identity associated with an increased proliferative potential. The same general mechanisms control epithelial regeneration in mammals (Fig. 2 and Table 1), although in this process a much larger orchestra of cell types and regulatory cytokines and growth factors is involved (Table 2), and these require more intricate conducting.

In addition to these classic inflammatory signalling pathways, the Wnt pathway — well known for its role in ASC homeostasis in mammals²⁰ — is emerging as an additional player in inflammatory tissue regeneration. Although it is less prominent than mammalian Wnt, Wingless signalling in flies drives self-renewal and cooperates with JAK–STAT to regulate homeostatic ISC proliferation and maintenance²¹. Strikingly, when flies sustain intestinal damage through ingestion of the bacterium *Pseudomonas entomophila* or the intestinal

irritant dextran sulfate sodium (DSS), Wntless expression is upregulated in enteroblasts — ISC daughter cells. These enteroblast-secreted, inflammatory Wnt signals activate downstream components that lead to enhanced ISC proliferation²². Similarly, inflammation-induced regeneration in mammals is guided by crosstalk between numerous cell types (such as the innate immune Paneth cells and ISCs that flank each other in crypts; Fig. 1a); the cytokines and growth factors that they produce; and ASCs (Fig. 2).

Notably, the constant contact that the gastrointestinal system has with microbes has allowed it to acquire important immune functions, including prevention of bacterial invasion and maintenance of tolerance. In addition, liver parenchymal cells, and to a lesser extent intestinal epithelial cells (IECs), metabolize and detoxify foodborne, waterborne and microbiota-generated toxic compounds. These protective functions are associated with a certain degree of collateral damage, which causes cell loss through physical attrition, chemical injury and immune destruction — processes that are especially pronounced in the mucosal lining of the gut. To prevent tissue loss and dysfunction and maintain homeostasis, the mammalian gastrointestinal system manifests strong regenerative capacity throughout its life.

Microbiota and regeneration

The human gut contains up to 100 trillion bacterial cells that belong to as many as a thousand different species, and a similar complexity is found in the murine gastrointestinal microbiota²³. The gut microbiota mainly consists of commensal microbes that exhibit symbiotic relationships with their host. In addition to modulating nutrient metabolism and absorption, the gut microbiota influences intestinal development and function²⁴, and shapes the gastrointestinal immune landscape²⁵. Mucosal erosion or injury allows commensal microbes and/or microbial macromolecules to penetrate and thereby activate macrophages, dendritic cells and T lymphocytes in the lamina propria. Activated immune cells produce numerous inflammatory cytokines, including tumour necrosis factor (TNF), IL-6, IL-10 and IL-17 family members. In addition to the propagation of intestinal inflammation, these cytokines control the regenerative response, which depends on ISC proliferation. Once the mucosa regenerates, microbial translocation and further inflammation are prevented. However, substantial disruption of the healthy microbiota caused by extensive and prolonged antibiotic use, especially in neonates and children, can result in a life-threatening pathology called necrotizing enterocolitis, in which mucosal injury results in cell death without regeneration²⁶. A similar condition can be elicited in mice by giving them broad-spectrum antibiotics, and can be prevented with oral administration of microbial products — such as lipopolysaccharide (LPS) — that induce the production of inflammatory cytokines through Toll-like receptor (TLR)4 activation²⁷. One such cytokine is IL-6, which prevents IEC death²⁸. However, TLR4 and IL-6 have also been implicated as pathogenic factors in necrotizing enterocolitis²⁹. Interestingly, although the intestinal microbiota of *Drosophila* is much simpler (featuring fewer than 20 microbial species), it also has an important role in intestinal regeneration. In the fly, symbiotic bacteria promote normal tissue growth, whereas potential pathogens produce uracil, which stimulates ROS production by activating a G-protein-coupled receptor (GPCR) that leads to p38 and JNK activation, induction of dual oxidase and ISC proliferation³⁰. Although limited ROS production stimulates ISC proliferation, extensive ROS generation can lead to the loss of epithelial homeostasis, which may underlie age-associated tissue dysfunction¹⁹. It remains to be determined whether a similar mechanism operates in the mammalian gut, in which the much more complex microbiome controls the amplitude of cytokine gene expression. More specific effects cannot be ruled out, and certain symbiotic microbes may even attenuate damaging inflammation.

The intestinal microbiota also influences liver regeneration³¹. The liver detoxifies LPS and other compounds derived from the gut microbiota that reach it through portal circulation. Thus, partial hepatectomy or extensive liver injury induced by carbon tetrachloride increases the

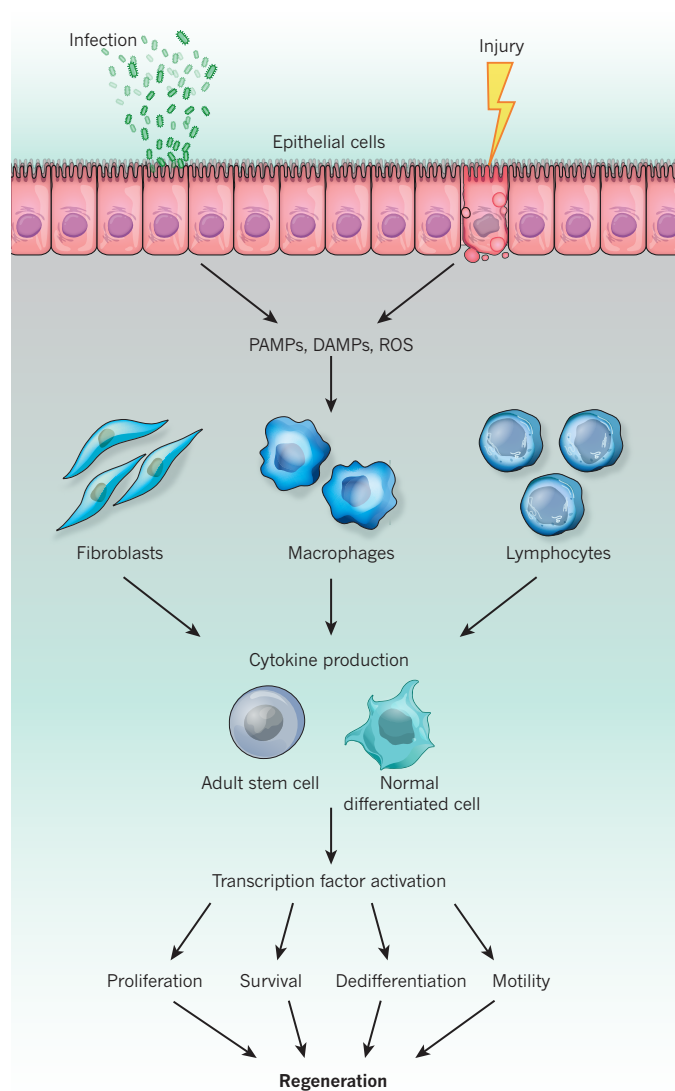


Figure 2 | Mechanisms through which infection and injury induce a regenerative inflammatory response. Injury or infection of epithelial tissues leads to the generation of pathogen-associated molecular patterns (PAMPs), damage-associated molecular patterns (DAMPs) and reactive oxygen species (ROS). These induce the production of cytokines (tumour necrosis factor, interleukin (IL)-6, IL-11, IL-17 or IL-22) through tissue constituents such as immune cells. Some of the most common inflammatory cytokines trigger signalling pathways in adult stem cells or normal differentiated cells that culminate in the activation of transcription factors (AP-1, NF- κ B, STAT3, YAP or Notch intracellular domain). These then mount a regenerative response by inducing genes that encode growth factors, stimulate cell-cycle progression, prevent cell death, promote dedifferentiation and the acquisition of ‘stemness’, and enhance cell motility and migration.

local concentration of LPS, which promotes liver regeneration through TLR4 activation and induction of inflammatory cytokines^{31,32}. Consistent with this hypothesis, germ-free rats have defective liver regeneration³¹, and administration of synbiotics (a mixture of probiotics and prebiotics) restores liver regeneration in these animals and enhances liver function in patients who have undergone hepatectomy³³. The gut microbiota also stimulates expression of reparative cytokines, partly through complement system activation^{34–36}.

Sterile inflammation and its sensors

In the absence of microbes, tissue damage and cell death still evoke sterile inflammation³⁷. As already mentioned, the microbiota mainly dictates the amplitude of the inflammatory response and its output, whereas damage-associated molecular patterns (DAMPs) initiate the

Table 1 | Sensors and activators of regenerative inflammation

Molecular class	Activators	Products
TLR	PAMPs and DAMPs	<i>Tnf</i> , <i>Il1β</i> , <i>Nlrp3</i> and <i>Mmp</i> mRNAs, IL-23, IL-6 and IL-22
NLR (NLRP3)	ATP, uric acid, mtDNA and mtROS	IL-1 β and IL-18
MMP/ADAM	TLR, TNF and IL-1 through AP-1 and NF- κ B	TNF, EGF, Areg and Ereg

Molecular class refers to the major sensors and mediators of regenerative inflammation. Areg, amphiregulin; DAMPs, damage-associated molecular patterns; EGF, epidermal growth factor; Ereg, epiregulin; IL, interleukin; MMP, matrix metalloprotease; mRNA, messenger RNA; mtDNA, mitochondrial DNA; mtROS, mitochondrial reactive oxygen species; NLR, NOD-like receptor; PAMPs, pathogen-associated molecular patterns; TLR, Toll-like receptor; TNF, tumour necrosis factor.

response³⁷. Cell death, especially necrotic death, and tissue damage cause the release of DAMPs, which include extracellular nucleic acids and chromatin components, ATP and other nucleotides, uric acid, cytoskeletal fragments, heat-shock proteins and oxidized mitochondrial DNA (mtDNA), all of which are sensed by PRRs. In addition to TLRs, many DAMPs affect membrane permeability to potassium and calcium, resulting in mitochondrial damage that culminates in leakage of mtDNA and mitochondrial ROS that cause inflammasome activation and IL-1 β and IL-18 secretion³⁸. The most important inflammasome for DAMP sensing is NLRP3, but its role in gastrointestinal regeneration and tissue repair is controversial. Some studies suggest that NLRP3 activation promotes epithelial integrity and regeneration, at least indirectly, through IL-1 and IL-18 production^{39,40}, but others have shown that IL-18 damages the intestinal mucosa, disrupts its barrier function⁴¹ and inhibits goblet-cell maturation during colitis⁴².

The response to sterile inflammatory triggers is evolutionarily conserved and has been studied in *Drosophila*, in which it is also involved in wound repair and regeneration¹⁷. In addition to PRR activation, sterile inflammation in flies and mammals results in activation of matrix metalloproteases (MMPs)^{43–45}, which are involved in the processing and release of cytokines and growth factors. The MMP ADAM17 controls regeneration of the injured colonic mucosa by shedding EGF receptor (EGFR) ligands and TNF⁴⁴. ADAM17 is also involved in Notch cleavage, providing a pathway through which inflammatory stimuli activate Notch signalling⁴⁵. In summary, TLR activation is responsible for the initial surge in cytokine gene transcription, inflammasome priming and MMP induction, whereas inflammasome activation controls IL-1 β and IL-18 secretion. MMP activation results in the release of cell-anchored cytokines and growth factors, such as TNF and EGF family members (Table 1).

Cytokines and tissue repair

Inflammation controls regenerative processes through several cytokines and growth factors. One of the first cytokines implicated in tissue repair was IL-6, which promotes liver regeneration after partial hepatectomy or CCl₄-induced injury^{34,35}. Shortly after, the primary inflammatory cytokine TNF was also found to control liver regeneration⁴⁶. Initially, these results were counter-intuitive because these cytokines, especially TNF, were thought to mediate tissue damage. But concurrent work revealed that TNF inhibits apoptotic cell death by activating NF- κ B^{47,48}. NF- κ B activation also inhibits necrosis⁴⁹, and a site-specific complement inhibitor — CR2–CD59, which blocks a membrane attack complex and increases hepatic TNF and IL-6 expression — strongly stimulates liver regeneration, even after 90% hepatectomy⁵⁰. IL-6 and TNF also promote regeneration of the injured intestinal mucosa^{28,51–54}, acting directly on epithelial cells by engaging IL-6 receptor (IL-6R):gp130 heterotetramers and TNF receptor 1 (TNFR1), respectively. TNF signalling also stimulates IL-6 expression, and TNF blockade in rats inhibits IL-6 production along with liver regeneration⁵⁵. TNFR1 engagement is required for stimulation of hepatocyte proliferation not only after partial hepatectomy⁵⁶ but also in hepatocellular carcinoma progenitors of mice with non-alcoholic steatohepatitis⁵⁷. TNF also promotes epithelial regeneration

through the Notch pathway⁵⁸, the activation of which may depend on ADAM17 and EGFR expression and engagement^{44,59}. Another TNF family member, lymphotoxin (LT), contributes to liver regeneration by binding to the LT β receptor⁶⁰. In *Drosophila*, TNF family members are used to control intestinal immunity and regeneration through the immune deficiency pathway¹⁹, further underscoring the ancient origin of inflammatory control of tissue repair.

IL-6 is the prototypical member of a large cytokine family, which also includes the mammalian proteins IL-11, IL-27, IL-31, leukaemia inhibitory factor (LIF), oncostatin M (OSM), leptin, ciliary neurotrophic factor (CNTF) and cardiotrophin-1, all of which are capable of stimulating cell proliferation and survival⁶¹. Most of these cytokines signal through heterodimeric or heterotetrameric receptors that use the gp130 signal transducing subunit⁶² (Fig. 3). IL-6-related proteins, known as Upd proteins, are present in *Drosophila*, in which they stimulate intestinal repair and regeneration either through direct effects on ISCs or by inducing expression of EGFR ligands in underlying muscle cells²¹. Other IL-6 family members, especially IL-11, LIF and OSM, may also contribute to gastrointestinal epithelial regeneration. In mammals, these cytokines can have indirect effects on epithelial regeneration by inducing the expression of EGFR ligands, the release of which is ADAM17-dependent^{44,63,64}. For instance, IL-6 induces amphiregulin (Areg) expression in the mammalian intestine⁶⁵.

Another important regenerative cytokine is IL-22, a member of the IL-10 family⁶⁶. IL-22 is produced by lymphocytes, especially T helper 17 (T_H17) cells and innate lymphoid cells (ILCs), and by certain myeloid subsets, but unlike most cytokines, it does not target other leukocytes. Instead, IL-22 acts on epithelial cells and fibroblasts to stimulate proliferation, inhibit death and delay terminal differentiation. IL-22 receptor engagement results in JAK–STAT3 activation, as well as mitogen-activated protein kinase (MAPK) activation, including extracellular signal-regulated kinase (ERK), p38 and JNK activation⁶⁶. Thus, IL-22 action is mainly dedicated to immune control of tissue repair. Fittingly, infection of IL-22-deficient mice with *Citrobacter rodentium* results in increased mucosal damage⁶⁷, and exogenous IL-22 ameliorates inflammation in a DSS-colitis model⁶⁸. IL-22 also prevents concanavalin-A-induced hepatocyte death^{69,70}, provides protection against acute pancreatitis by inducing Reg family members⁶⁶ and can directly stimulate proliferation of epithelial cells, including ISCs⁶⁸. Like T_H17 cells, ILCs also produce IL-17 cytokines, which include the six members A to F. These cytokines are important regulators of epithelial barrier integrity⁷¹. Curiously, the primordial member of this family is the *Drosophila* protein Spätzle, the ligand of Toll (or Toll-1, the prototypical TLR)⁷². IL-17 cytokines bind heterodimeric receptors that use IL-17RA as a common subunit⁷³, the engagement of which causes activation of NF- κ B and MAPKs, which in turn induce antimicrobial immunity and tissue remodelling. IL-17-induced genes encode antimicrobial peptides, such as β -defensins and Reg3y, and inflammatory cytokines, including IL-6 (refs 71, 74). Like IL-22 deficient mice, IL-17RE-deficient mice exhibit enhanced mucosal damage after *C. rodentium* infection⁷⁵. Although this phenotype may be due to the antimicrobial effects of IL-17 signalling, IL-17RA engagement can directly stimulate IEC proliferation⁷⁶. Curiously, IL-17 cytokines are not as important in liver regeneration as IL-22 (ref. 69); the stronger regenerative effect of IL-22 could be due to its ability to directly activate STAT3. IL-17 cytokines, however, can lead to indirect STAT3 activation in epithelial cells by stimulating immune cells to produce IL-6 (ref. 76). See Table 2 for a summary of key regeneration-stimulating cytokines.

Signalling pathways in inflammation-led regeneration

Several evolutionarily conserved signalling pathways connect inflammatory inputs to the regenerative response (Fig. 2). First and foremost is the MAPK–AP-1 pathway, which also contributes to inflammation-driven regeneration and repair in *Drosophila*^{16,18}. Even in *Drosophila*, JNK and p38 have additional targets, including Foxo, which controls stress resistance and antioxidant gene expression, and the AP-1-related

factor ATF2, which controls dual-oxidase expression¹⁶. In mammals, the AP-1 component c-Jun controls liver regeneration, partly by suppressing p53 and p38 MAPK activities⁷⁷ and inducing cyclin D1 expression on JNK1 activation⁷⁸. JNK–AP-1 signalling, however, can have opposing and complex effects on hepatocyte proliferation and survival, partly through the inhibition of pro-survival NF- κ B activity^{79–81}, the induction of Nos2-generated pro-survival signals⁸², the stimulation of cell-cycle progression⁷⁸ and the modulation of mitochondria-dependent apoptosis⁸³. JNK-dependent AP-1 and p38 MAPK also act in non-parenchymal liver cells to induce TNF and IL-6, which, as already described, control liver regeneration⁸³. EGFR ligands regulate epithelial homeostasis through MAPKs in the *Drosophila* midgut⁸⁴ and the mammalian intestine, in which ERK MAPKs control ISC proliferation and migration of their progeny along the crypt–villus axis. Src-mediated p38 activation stimulates the migration, but not the proliferation, of an IEC-derived cell line⁸⁵, but this remains to be demonstrated *in vivo*. p38 and other MAPKs also exert regenerative effects through ADAM17 activation⁴⁵.

Another evolutionarily conserved inflammatory signalling pathway that controls epithelial tissue integrity and survival is the IKK–NF- κ B pathway⁸⁶. NF- κ B activation is needed to protect hepatocytes from TNF-induced apoptosis and allow them to respond to proliferative signals generated by TNFR1 engagement⁸⁷. Some of NF- κ B's protective effects are mediated through GADD45 β , which inhibits prolonged JNK activation. Others, however, depend on inducible anti-apoptotic proteins such as c-Flip^{79,80}. NF- κ B activation in immune cells contributes to liver regeneration and hepatocyte proliferation by promoting synthesis of IL-6, TNF, LT and other cytokines⁸⁸. IKK β -dependent NF- κ B also protects the intestinal epithelium from injury induced by DSS, ionizing radiation or ischaemia–reperfusion^{51,89–91}. Epithelial IKK β , however, has no protective role in chronic colitis caused by IL-10 deficiency⁹². NF- κ B may further enhance proliferation of differentiated IECs by potentiating β -catenin signalling⁹³, and like AP-1 it contributes to the induction of IL-6 and other cytokines in lamina propria immune cells⁸⁹. These cytokines promote IEC survival and ISC proliferation by activating STAT3 (ref. 52) (Fig. 3); the *Drosophila* counterpart DStat responds to Upd proteins, which control ISC proliferation⁹⁴. STAT3 is activated by the tyrosine kinases JAK1 and JAK2 in response to gp130 or IL-22R signalling and is subject to feedback inhibition by SOCS3 (ref. 95). All of these regulators positively or negatively modulate intestinal homeostasis⁹⁵ and are present in *Drosophila*, in which they are known as Hop (DJAK) and Socs (SOCS3). STAT3 also contributes to expression of antimicrobial peptides in Paneth cells^{96,97}, which provide a niche for mammalian ISCs. Like NF- κ B, with which it collaborates⁹⁸, STAT3 is not needed for the development or maintenance of the uninjured intestinal epithelium, possibly because its loss may be compensated for by STAT1 (ref. 99). Notably, human STAT3 and JAK2 genes have been identified as susceptibility loci for inflammatory bowel disease¹⁰⁰. STAT3 is activated by IL-6 and related cytokines, and is required for the stimulation of liver regeneration after partial hepatectomy, toxic damage or inflammation-induced injury¹⁰¹ because it induces genes with products that maintain cell survival and promote proliferation¹⁰².

Two other functionally linked and evolutionarily conserved signalling pathways with key roles in regeneration and gastrointestinal homeostasis are the Hippo–YAP (Fig. 3) and Notch signalling pathways^{94,103,104}. YAP and its orthologue TAZ are transcriptional co-activators of TEAD transcription factors, which control a gene-expression program that stimulates cell proliferation, suppresses cell death and induces other receptors and ligands^{103,105–107}. In non-stimulated epithelial cells, YAP and TAZ remain in the cytoplasm and undergo proteasomal degradation, as a result of phosphorylation of inhibitory serine residues by Warts (in *Drosophila*) or LATS1/2 kinases (in mammals)¹⁰⁷. Genetic loss of these protein kinases or their activators (Hippo or MST1/2) results in YAP/TAZ dephosphorylation, stabilization, nuclear translocation and activation of TEAD and other transcription factors, with which these co-activators interact. Loss of cell–cell contact, cell adhesion or cytoskeletal integrity leads to transient Hippo/MST inhibition and

Table 2 | Regeneration-inducing cytokines and their properties

Cytokines	Source	Direct effectors
TNF	T lymphocytes, macrophages and epithelial cells	NF- κ B, MAPKs and AP-1
IL-6	Lymphocytes, myeloid cells, fibroblasts and epithelial cells	JAK–STAT3, MAPKs, SFKs, YAP and Notch
IL-22	T _H 17 cells, iLCs and some myeloid cells	JAK–STAT3, MAPKs, SFKs, YAP and Notch
IL-17	T _H 17 cells, $\gamma\delta$ T cells and iLCs	NF- κ B and MAPKs

IL, interleukin; iLCs, innate lymphoid cells; MAPKs, mitogen-activated protein kinases; SFKs, Src family kinases; T_H17, T helper 17; TNF, tumour necrosis factor.

consequent YAP/TAZ activation¹⁰⁷. In both *Drosophila* and mammals, the Hippo pathway restricts uncontrolled ISC proliferation, whereas activation of YAP and its fly homologue Yorkie is required for intestinal epithelial regeneration after injury^{108–114}. More refined genetic analysis in *Drosophila* demonstrated that Hippo–YAP signalling in differentiated IECs controls ISC proliferation, such that Yorkie/YAP activation after epithelial injury results in the production of signals by differentiated cells that act on neighbouring ISCs. Given the dependence of inhibitory Hippo signalling on cell–cell and cell–matrix contact, or cytoskeletal integrity, it was assumed that disruption of the gut epithelial lining results in Yorkie/YAP activation owing to inhibition of Hippo/MST kinase activity. However, an entirely different mechanism has recently been identified, at least in mammals (Fig. 3). Injury of the intestinal mucosa results in penetration of commensal microbes or their products (for example, LPS), leading to induction of IL-6, IL-11, IL-22 and related cytokines by lamina propria macrophages and dendritic cells. IL-6 and IL-11 activate gp130 signalling and induce the association of tyrosine-phosphorylated gp130 with the Src family kinases (SFJs) Src and Yes, thereby stimulating their tyrosine kinase activity⁶⁵. Activated SFJs interact with YAP and phosphorylate it to induce its stabilization and nuclear translocation⁶⁵. YAP target genes induced through this pathway include those that encode Notch receptors and ligands, and Areg, which acts on ISCs to enhance their proliferation. These findings demonstrate that YAP is also subject to positive regulation and can be rapidly activated in response to inflammatory signals, which, as well as IL-6 family members, may also include IL-22. Accordingly, interference with SFJ-dependent YAP activation attenuates inflammation-induced intestinal regeneration even though it does not affect the parallel JAK–STAT3 pathway⁶⁵.

YAP activation in IECs induces several Notch receptors and ligands, including Notch1, Notch3 and DLL3 (ref. 65). This results in activation of Notch, which maintains ISCs and transit-amplifying cells in the crypt compartment in a highly proliferative and undifferentiated state^{115,116}. Persistent Notch activation inhibits the generation of secretory cell types (goblet, enteroendocrine and Paneth cells) and slows the differentiation of absorptive enterocytes^{115,117}. A similar phenotype has been observed with persistent YAP activation owing to either MST1/2 ablation^{112,113} or intestinal-specific expression of a constitutively active gp130 variant⁶⁵. γ -Secretase inhibition in mice with activated Notch or gp130 signalling in IECs led to the restoration of epithelial differentiation and homeostasis^{65,115}. Conversely, ectopic expression of activated Notch in the *Drosophila* gut results in rapid differentiation of ISCs into enterocytes and inhibition of secretory-cell formation, along with ISC depletion¹¹⁸. Of note, *Drosophila* Src is also involved in intestinal regeneration, although its downstream targets have not been identified¹¹⁹.

The Wnt signalling cascade is arguably the most prominent regulator of ASCs in mammalian epithelia. Its role was first revealed in the maintenance of small intestinal crypt stem cells¹²⁰. It has since been extended to other perpetually self-renewing tissues, including hair follicles, the colon and stomach epithelium²⁰. Most other organs have very little homeostatic proliferation under physiological conditions, but respond

with a burst of regenerative proliferation when damaged. Although not extensively investigated yet, it seems likely that Wnts have a central role in regenerative stem-cell activity. The Wnt target gene products *Lgr5* and *Axin2* have emerged as common markers of constitutive, as well as damage-activated, types of Wnt-driven ASCs²⁰ (Fig. 3).

Despite intense ISC baseline activity, intestinal crypts become even more active (hyperplastic) during episodes of damage and inflammation, further increasing ISC and Paneth cell numbers and cellular output. Indeed, mice carrying a loss-of-function allele of the Wnt-agonistic receptor *Lgr4* have a considerably higher susceptibility to DSS-induced colitis, concomitant with greatly reduced ISC numbers¹²¹. During these temporary hyperplastic phases, Wnt signalling activity increases markedly¹²². In the colon, regeneration of lost epithelium and the subsequent correct patterning of new crypts involves non-canonical Wnt5a activity¹²³. Increases in size and number of Paneth cells are typically seen during an inflammatory or damage response, and consequent crypt hyperplasia can be viewed as a special form of inflammation that is not mediated by bone-marrow-derived immune cells, but by an endodermal epithelial cell with a prominent role in host defence and regeneration — the Paneth cell¹²⁴. Paneth cells that are normally restricted to the small bowel appear in the human colon alongside chronic inflammation, a phenomenon that is well known to pathologists and called Paneth cell metaplasia. But other regenerative signalling molecules, such as Notch and YAP, inhibit Paneth-cell formation^{113,115}. Thus, the size of the ISC niche is probably kept in check by the balance between the different regeneration-promoting pathways reviewed above. Since the original

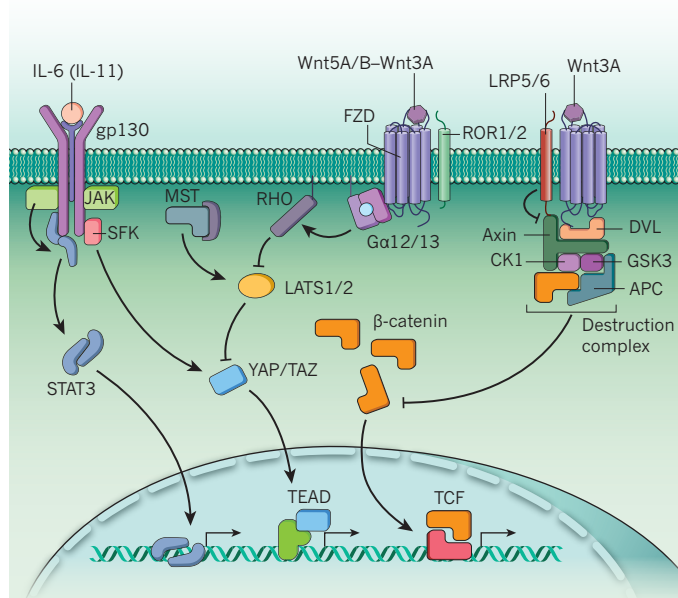


Figure 3 | Signalling pathways in inflammation-driven regeneration.

Cytokine signalling, the Wnt pathway (canonical and non-canonical) and Hippo-independent Src family kinase (SFK)–YAP signalling connects inflammatory signals to the regenerative response. Interleukin (IL)-6, and other members of the cytokine family, signal through receptors that use the gp130 subunit. STAT3 is activated by the tyrosine kinases JAK1 and JAK2 in response to gp130 or IL-22R signalling. In addition, cytokine signalling recruits SFK to gp130, which leads to phosphorylation, stabilization and nuclear translocation of YAP and its orthologue TAZ. Nuclear YAP and TAZ act as transcriptional co-activators of TEAD transcription factors. In non-stimulated epithelial cells, YAP and TAZ remain in the cytoplasm and undergo proteasomal degradation as a result of phosphorylation of inhibitory serine residues by LATS1/2 kinases. LATS1/2 is activated by Hippo/MST and inhibited by non-canonical Wnt signalling (FZD–ROR1/2 complex) owing to Ga12/13-dependent Rho activation. Canonical Wnt signalling (FZD–LRP5/6 complex), however, stabilizes β -catenin by inhibiting its continuous degradation by the destruction complex. Stabilized β -catenin can then translocate to the nucleus and stimulate TCF-mediated gene transcription.

description of label-retaining ISCs, or +4 cells, near the crypt bottom¹²⁵, these cells — which serve as a reserve for the ISC population — have been the focus of intense research. The most notable marker for the +4 cell has been *Bmi1* (ref. 126). Although +4 cell markers such as *Bmi1* seem to be shared between *Lgr5*⁺ and +4 ISCs^{127,128}, a pool of transient, non-dividing Paneth cell precursors near the crypt base may actually serve as the reserve stem-cell pool during periods of damage^{128,129}. The signals that control these quiescent cells have not been revealed, but it seems safe to assume that Wnt signals have a central role in this process.

Wnt signals are also emerging as crucial regulators of liver repair. A study in the zebrafish *Danio rerio* showed that biliary epithelial cells (BECs) are a crucial stem-cell source for injury repair when hepatocyte proliferation is compromised. Massive hepatocyte loss results in the dedifferentiation of BECs into hepatoblast-like cells and the subsequent formation of highly proliferative hepatocytes to restore liver mass. This process is strongly dependent on an intact *Wnt2b* gene, implying that Wnt signals are involved in this oval-cell-like response¹³⁰. Nonetheless, lineage-tracing studies in mice have failed to detect any contribution from oval cells or other bile-duct-derived cells to hepatocyte regeneration after chemical injury¹³¹. Furthermore, a population of periportal hepatocytes that do not express the metabolic functions that characterize fully differentiated hepatocytes have been identified in the normal liver, surrounding the central vein (Fig. 1b). These cells — termed hybrid hepatocytes because they express *Sox9* and other bile-duct genes — do not metabolize CCl_4 (and therefore escape its toxic effect), undergo several rounds of proliferation and repair the damage¹³². When transplanted into a diseased liver, hybrid hepatocytes have higher regenerative capacity than normal hepatocytes or oval cells.

Although Wnts may control the weak proliferative activity of diploid pericentral hepatocytes, Wnt signalling seems to be particularly important for ‘waking up’ quiescent cells in the biliary tree after generalized liver damage^{133–135}. The Wnt-dependent stem-cell marker *Lgr5* is not expressed in healthy liver, but it is induced in small cells that carry biliary markers when mice are given hepatocyte toxins, such as CCl_4 (ref. 136). Indeed, CCl_4 treatment leads to a massive induction of Wnt proteins and Wnt-supporting R-spondins in the damaged liver. Lineage tracing has revealed that these induced *Lgr5*⁺ cells generate large numbers of hepatocytes and bile-duct cells in the damaged areas, which is indicative of their bipotency. Such *Lgr5*⁺ cells can be grown over long periods of time in a Wnt-based three-dimensional culture system as epithelial organoids that contain hepatocyte-like cells and BECs. Similarly, a human BEC generates ever-growing organoids that consist of hepatocyte-like cells as well as BECs. When transplanted into mice, BECs yield mature hepatocytes¹³⁷. The source of Wnts and R-spondins during CCl_4 -induced damage remains to be determined, but it has been reported that the resident liver macrophages (Kupffer cells) are a major source of Wnt after partial hepatectomy⁴. Similarly, in a model of chronic liver damage, macrophage engulfment of sterile hepatocyte debris induced Wnt3a expression, and when these macrophages were removed, no new hepatocytes were formed¹³⁸. Thus, although more studies are required, it seems that Wnts produced by inflammatory cells in the damaged liver and Wnt-pathway activity are essential components of the signalling toolbox that the liver exploits for its regeneration. The relationship between hybrid hepatocytes and the *Lgr5*⁺ progenitors previously discussed also needs to be investigated. Finally, an alternative mode of Wnt signalling has been described that involves the non-canonical Wnt receptor–ligand pair ROR1/2 and Wnt5a, which directly activates YAP signalling¹³⁹ (Fig. 3). This provides a further example of the close interconnection between regenerative signalling pathways.

Harnessing inflammatory regeneration for therapy

Balancing the positive and negative effects of the inflammatory reaction has been key for the design of clinical treatments for a multitude of diseases. The advantages of removing harmful foreign agents and infected or damaged cells need to be weighed against the disadvantages of chronic or uncontrolled inflammation in which friendly

inflammatory fire causes more harm than good. In this Review, we have discussed an aspect of inflammation that has received much less attention. The inflammatory reaction not only deals with what has to be removed but also supports the rebuilding of what has been lost. In other words, it is an important driver of the regenerative response, an ancient function that has been evolutionarily conserved between flies and mammals. Orchestrating the action of the numerous cellular components involved in sensing and executing inflammation, and balancing the ensuing regenerative response are key to optimal recovery. Indeed, regeneration and mucosal healing have been suggested as key treatment goals that predict sustained remission and resection-free survival for inflammatory bowel disease²⁸. But the key question is this: which of the known regenerative cytokines or signalling molecules can be harnessed to achieve this therapeutic goal? Answering this question is fundamental, because when regeneration falls short, there will be insufficient tissue for the affected organ to function, and when regeneration goes awry, the wrong tissue might be produced, resulting in scar formation or fibrosis. Although myofibroblasts are essential components of both inflammatory and regenerative responses, their extensive proliferation without subsequent removal causes aberrant regeneration and excessive collagen deposition, all of which fall under the fibrosis umbrella¹⁴. Worse still, the signalling pathways that are activated during inflammation to support regeneration are, in one way or another, all drivers of cancer — a known complication of chronic inflammation¹³. Furthermore, regenerative cytokines such as IL-17 or IL-22 may be responsible for the development of resistance to cancer drugs¹⁴⁰. However, the neutralization of these regenerative responses creates another serious problem. The major dose-limiting factor in chemotherapy or radiotherapy for cancer is mucositis — severe mucosal inflammation. Blockade of cancer-promoting inflammatory signals should not, therefore, be combined with mucositis-inducing therapies. Likewise, maintenance of liver function is essential for proper detoxification of chemotherapeutic drugs, and the presence of liver fibrosis prevents the use of such drugs in the treatment of liver cancer. Obviously, much remains to be learned about the role of inflammation in functional and practical tissue restoration. Undoubtedly, such knowledge will be crucial to support optimal tissue repair — the right kind of tissue produced in the right amount and at the right location — during an ongoing inflammatory response.

Note added in proof: A paper recently appeared online while the current Review was in press that underscored the regenerative role of IL-22 by demonstrating that it directly promotes proliferation of isolated ISCs in culture (C. A. Lindemans *et al.* Interleukin-22 promotes intestinal-stem-cell-mediated epithelial regeneration *Nature* <http://dx.doi.org/10.1038/nature16460>; 2015). ■

Received 11 August; accepted 7 December 2015.

- Clevers, H. The intestinal crypt, a prototype stem cell compartment. *Cell* **154**, 274–284 (2013).
- Sato, T. *et al.* Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469**, 415–418 (2011).
- This paper outlines how Paneth cells provide support for ISCs.**
- Durand, A. *et al.* Functional intestinal stem cells after Paneth cell ablation induced by the loss of transcription factor Math1 (Atoh1). *Proc. Natl Acad. Sci. USA* **109**, 8965–8970 (2012).
- Yang, J. *et al.* β -catenin signaling in murine liver zonation and regeneration: a Wnt–Wnt situation! *Hepatology* **60**, 964–976 (2014).
- Wang, B., Zhao, L., Fish, M., Logan, C. Y. & Nusse, R. Self-renewing diploid Axin2⁺ cells fuel homeostatic renewal of the liver. *Nature* **524**, 180–185 (2015).
- This paper describes a population of diploid pericentral hepatocytes that may act as adult liver stem cells.**
- Stanger, B. Z. Cellular homeostasis and repair in the mammalian liver. *Annu. Rev. Physiol.* **77**, 179–200 (2015).
- Sun, G. & Irvine, K. D. Control of growth during regeneration. *Curr. Top. Dev. Biol.* **108**, 95–120 (2014).
- Monga, S. P. Role and regulation of β -catenin signaling during physiological liver growth. *Gene Expr.* **16**, 51–62 (2014).
- Hu, M. *et al.* Wnt/ β -catenin signaling in murine hepatic transit amplifying progenitor cells. *Gastroenterology* **133**, 1579–1591 (2007).
- Farber, E. Similarities in the sequence of early histological changes induced in the liver of the rat by ethionine, 2-acetylaminofluorene, and 3'-methyl-4-dimethylaminoazobenzene. *Cancer Res.* **16**, 142–148 (1956).

- Popper, H., Kent, G. & Stein, R. Ductular cell reaction in the liver in hepatic injury. *J. Mt. Sinai Hosp.* **24**, 551–556 (1957).
- West, N. R., McCuaig, S., Franchini, F. & Powrie, F. Emerging cytokine networks in colorectal cancer. *Nature Rev. Immunol.* **15**, 615–629 (2015).
- Grivennikov, S. I., Greten, F. R. & Karin, M. Immunity, inflammation, and cancer. *Cell* **140**, 883–899 (2010).
- Stramer, B. M., Mori, R. & Martin, P. The inflammation-fibrosis link? A Jekyll and Hyde role for blood cells during wound repair. *J. Invest. Dermatol.* **127**, 1009–1017 (2007).
- Gurtner, G. C., Werner, S., Barrandon, Y. & Longaker, M. T. Wound repair and regeneration. *Nature* **453**, 314–321 (2008).
- Panayidou, S. & Apidianakis, Y. Regenerative inflammation: lessons from *Drosophila* intestinal epithelium in health and disease. *Pathogens* **2**, 209–231 (2013).
- Shaukat, Z., Liu, D. & Gregory, S. Sterile inflammation in *Drosophila*. *Mediators Inflamm.* **2015**, 369286 (2015).
- Buchon, N., Silverman, N. & Cherry, S. Immunity in *Drosophila melanogaster* — from microbial recognition to whole-organism physiology. *Nature Rev. Immunol.* **14**, 796–810 (2014).
- Ayyaz, A. & Jasper, H. Intestinal inflammation and stem cell homeostasis in aging *Drosophila melanogaster*. *Front. Cell Infect. Microbiol.* **3**, 98 (2013).
- Clevers, H., Loh, K. M. & Nusse, R. An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control. *Science* **346**, 1248012 (2014).
- Xu, N. *et al.* EGFR, Wntless and JAK/STAT signaling cooperatively maintain *Drosophila* intestinal stem cells. *Dev. Biol.* **354**, 31–43 (2011).
- Cordero, J. B., Stefanatos, R. K., Scopelliti, A., Vidal, M. & Sansom, O. J. Inducible progenitor-derived Wntless regulates adult midgut regeneration in *Drosophila*. *EMBO J.* **31**, 3901–3917 (2012).
- Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489**, 220–230 (2012).
- Sommer, F. & Backhed, F. The gut microbiota — masters of host development and physiology. *Nature Rev. Microbiol.* **11**, 227–238 (2013).
- Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions between the microbiota and the immune system. *Science* **336**, 1268–1273 (2012).
- Neu, J. & Walker, W. A. Necrotizing enterocolitis. *N. Engl. J. Med.* **364**, 255–264 (2011).
- Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S. & Medzhitov, R. Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell* **118**, 229–241 (2004).
- This is one of the first reports to describe the role of TLR4 signalling in control of mucosal homeostasis.**
- Neurath, M. F. New targets for mucosal healing and therapy in inflammatory bowel diseases. *Mucosal Immunol.* **7**, 6–19 (2014).
- Claud, E. C. Neonatal necrotizing enterocolitis — inflammation and intestinal immaturity. *Antiinflamm. Antiallergy Agents Med. Chem.* **8**, 248–259 (2009).
- Lee, W. J. & Brey, P. T. How microbiomes influence metazoan development: insights from history and *Drosophila* modeling of gut-microbe interactions. *Annu. Rev. Cell Dev. Biol.* **29**, 571–592 (2013).
- Cornell, R. P., Liljequist, B. L. & Bartzall, K. F. Depressed liver regeneration after partial hepatectomy of germ-free, athymic and lipopolysaccharide-resistant mice. *Hepatology* **11**, 916–922 (1990).
- Seki, E. *et al.* Contribution of Toll-like receptor/myeloid differentiation factor 88 signaling to murine liver regeneration. *Hepatology* **41**, 443–450 (2005).
- This is one of the first accounts of the control of liver regeneration by TLR signalling.**
- Rayes, N. *et al.* Effect of pre- and probiotics on liver regeneration after resection: a randomised, double-blind pilot study. *Benef. Microbes* **3**, 237–244 (2012).
- Taub, R. Liver regeneration: from myth to mechanism. *Nature Rev. Mol. Cell Biol.* **5**, 836–847 (2004).
- Cressman, D. E. *et al.* Liver failure and defective hepatocyte regeneration in interleukin-6-deficient mice. *Science* **274**, 1379–1383 (1996).
- This key paper describes the regenerative function of IL-6.**
- DeAngelis, R. A. *et al.* A complement-IL-4 regulatory circuit controls liver regeneration. *J. Immunol.* **188**, 641–648 (2012).
- Rock, K. L., Latz, E., Ontiveros, F. & Kono, H. The sterile inflammatory response. *Annu. Rev. Immunol.* **28**, 321–342 (2010).
- Elliott, E. I. & Sutterwala, F. S. Initiation and perpetuation of NLRP3 inflammasome activation and assembly. *Immunol. Rev.* **265**, 35–52 (2015).
- Dupaul-Chicoine, J. *et al.* Control of intestinal homeostasis, colitis, and colitis-associated colorectal cancer by the inflammatory caspases. *Immunity* **32**, 367–378 (2010).
- Zaki, M. H. *et al.* The NLRP3 inflammasome protects against loss of epithelial integrity and mortality during experimental colitis. *Immunity* **32**, 379–391 (2010).
- Lopetuso, L. R., Chowdhry, S. & Pizarro, T. T. Opposing functions of classic and novel IL-1 family members in gut health and disease. *Front. Immunol.* **4**, 181 (2013).
- Nowarski, R. *et al.* Epithelial IL-18 equilibrium controls barrier function in colitis. *Cell* **163**, 1–13 (2015).
- Stevens, L. J. & Page-McCaw, A. A secreted MMP is required for re-epithelialization during wound healing. *Mol. Biol. Cell* **23**, 1068–1079 (2012).
- Chalaris, A. *et al.* Critical role of the disintegrin metalloprotease ADAM17 for intestinal inflammation and regeneration in mice. *J. Exp. Med.* **207**, 1617–1624 (2010).

45. Scheller, J., Chalaris, A., Garbers, C. & Rose-John, S. ADAM17: a molecular switch to control inflammation and tissue regeneration. *Trends Immunol.* **32**, 380–387 (2011).
46. Yamada, Y., Kirillova, I., Peschon, J. J. & Fausto, N. Initiation of liver growth by tumor necrosis factor: deficient liver regeneration in mice lacking type I tumor necrosis factor receptor. *Proc. Natl Acad. Sci. USA* **94**, 1441–1446 (1997).
47. Liu, Z. G., Hsu, H., Goeddel, D. V. & Karin, M. Dissection of TNF receptor 1 effector functions: JNK activation is not linked to apoptosis while NF- κ B activation prevents cell death. *Cell* **87**, 565–576 (1996).
48. Beg, A. A. & Baltimore, D. An essential role for NF- κ B in preventing TNF- α -induced cell death. *Science* **274**, 782–784 (1996).
49. Kamata, H. *et al.* Reactive oxygen species promote TNF α -induced death and sustained JNK activation by inhibiting MAP kinase phosphatases. *Cell* **120**, 649–661 (2005).
50. Marshall, K. M., He, S., Zhong, Z., Atkinson, C. & Tomlinson, S. Dissecting the complement pathway in hepatic injury and regeneration with a novel protective strategy. *J. Exp. Med.* **211**, 1793–1805 (2014).
51. Chen, L. W. *et al.* The two faces of IKK and NF- κ B inhibition: prevention of systemic inflammation but increased local injury following intestinal ischemia-reperfusion. *Nature Med.* **9**, 575–581 (2003).
This is the first account of the crucial protective and regenerative function of TNF-induced intestinal NF- κ B.
52. Grivennikov, S. *et al.* IL-6 and Stat3 are required for survival of intestinal epithelial cells and development of colitis-associated cancer. *Cancer Cell* **15**, 103–113 (2009).
53. Becker, C. *et al.* TGF- β suppresses tumor progression in colon cancer by inhibition of IL-6 trans-signaling. *Immunity* **21**, 491–501 (2004).
54. Böhm, F., Kohler, U. A., Speicher, T. & Werner, S. Regulation of liver regeneration by growth factors and cytokines. *EMBO Mol. Med.* **2**, 294–305 (2010).
55. Akerman, P. *et al.* Antibodies to tumor necrosis factor- α inhibit liver regeneration after partial hepatectomy. *Am. J. Physiol.* **263**, G579–G585 (1992).
56. Yamada, Y., Webber, E. M., Kirillova, I., Peschon, J. J. & Fausto, N. Analysis of liver regeneration in mice lacking type 1 or type 2 tumor necrosis factor receptor: requirement for type 1 but not type 2 receptor. *Hepatology* **28**, 959–970 (1998).
57. Nakagawa, H. *et al.* ER stress cooperates with hypernutrition to trigger TNF-dependent spontaneous HCC development. *Cancer Cell* **26**, 331–343 (2014).
58. Ando, K. *et al.* Induction of Notch signaling by tumor necrosis factor in rheumatoid synovial fibroblasts. *Oncogene* **22**, 7796–7803 (2003).
59. Hilliard, V. C., Frey, M. R., Dempsey, P. J., Peek, R. M. Jr & Polk, D. B. TNF- α converting enzyme-mediated ErbB4 transactivation by TNF promotes colonic epithelial cell survival. *Am. J. Physiol. Gastrointest. Liver Physiol.* **301**, G338–G346 (2011).
60. Anders, R. A., Subudhi, S. K., Wang, J., Pfeffer, K. & Fu, Y. X. Contribution of the lymphotoxin β receptor to liver regeneration. *J. Immunol.* **175**, 1295–1300 (2005).
61. Garbers, C. *et al.* Plasticity and cross-talk of interleukin 6-type cytokines. *Cytokine Growth Factor Rev.* **23**, 85–97 (2012).
62. Kishimoto, T. IL-6: from its discovery to clinical applications. *Int. Immunol.* **22**, 347–352 (2010).
63. Brandl, K. *et al.* MyD88 signaling in non-hematopoietic cells protects mice against induced colitis by regulating specific EGF receptor ligands. *Proc. Natl Acad. Sci. USA* **107**, 19967–19972 (2010).
64. Makki, N., Thiel, K. W. & Miller, F. J. Jr. The epidermal growth factor receptor and its ligands in cardiovascular disease. *Int. J. Mol. Sci.* **14**, 20597–20613 (2013).
65. Taniguchi, K. *et al.* A gp130–Src–YAP module links inflammation to epithelial regeneration. *Nature* **519**, 57–62 (2015).
This key paper describes the role of gp130-induced, Hippo-independent YAP signalling in epithelial regeneration.
66. Nikoipour, E., Bellemore, S. M. & Singh, B. IL-22, cell regeneration and autoimmunity. *Cytokine* **74**, 35–42 (2015).
67. Zheng, Y. *et al.* Interleukin-22 mediates early host defense against attaching and effacing bacterial pathogens. *Nature Med.* **14**, 282–289 (2008).
68. Sugimoto, K. *et al.* IL-22 ameliorates intestinal inflammation in a mouse model of ulcerative colitis. *J. Clin. Invest.* **118**, 534–544 (2008).
69. Zenewicz, L. A. *et al.* Interleukin-22 but not interleukin-17 provides protection to hepatocytes during acute liver inflammation. *Immunity* **27**, 647–659 (2007).
This paper is an important account of the unique regenerative function of IL-22.
70. Radaeva, S., Sun, R., Pan, H. N., Hong, F. & Gao, B. Interleukin 22 (IL-22) plays a protective role in T cell-mediated murine hepatitis: IL-22 is a survival factor for hepatocytes via STAT3 activation. *Hepatology* **39**, 1332–1342 (2004).
71. Pappu, R., Rutz, S. & Ouyang, W. Regulation of epithelial immunity by IL-17 family cytokines. *Trends Immunol.* **33**, 343–349 (2012).
72. Hymowitz, S. G. *et al.* IL-17s adopt a cystine knot fold: structure and activity of a novel cytokine, IL-17F, and implications for receptor binding. *EMBO J.* **20**, 5332–5341 (2001).
73. Ely, L. K., Fischer, S. & Garcia, K. C. Structural basis of receptor sharing by interleukin 17 cytokines. *Nature Immunol.* **10**, 1245–1251 (2009).
74. Iwakura, Y., Ishigame, H., Saijo, S. & Nakae, S. Functional specialization of interleukin-17 family members. *Immunity* **34**, 149–162 (2011).
75. Song, X. *et al.* IL-17RE is the functional receptor for IL-17C and mediates mucosal immunity to infection with intestinal pathogens. *Nature Immunol.* **12**, 1151–1158 (2011).
76. Wang, K. *et al.* Interleukin-17 receptor a signaling in transformed enterocytes promotes early colorectal tumorigenesis. *Immunity* **41**, 1052–1063 (2014).
77. Stepniak, E. *et al.* c-Jun/AP-1 controls liver regeneration by repressing p53/p21 and p38 MAPK activity. *Genes Dev.* **20**, 2306–2314 (2006).
78. Sakurai, T., Maeda, S., Chang, L. & Karin, M. Loss of hepatic NF- κ B activity enhances chemical hepatocarcinogenesis through sustained c-Jun N-terminal kinase 1 activation. *Proc. Natl Acad. Sci. USA* **103**, 10544–10551 (2006).
79. Chang, L. *et al.* The E3 ubiquitin ligase itch couples JNK activation to TNF α -induced cell death by inducing c-FLIP₁ turnover. *Cell* **124**, 601–613 (2006).
80. Papa, S. *et al.* Gadd45 β promotes hepatocyte survival during liver regeneration in mice by modulating JNK signaling. *J. Clin. Invest.* **118**, 1911–1923 (2008).
81. Schwabe, R. F. & Brenner, D. A. Mechanisms of liver injury. I. TNF- α -induced liver injury: role of IKK, JNK, and ROS pathways. *Am. J. Physiol. Gastrointest. Liver Physiol.* **290**, G583–G589 (2006).
82. Hasselblatt, P., Rath, M., Komnenovic, V., Zatloukal, K. & Wagner, E. F. Hepatocyte survival in acute hepatitis is due to c-Jun/AP-1-dependent expression of inducible nitric oxide synthase. *Proc. Natl Acad. Sci. USA* **104**, 17105–17110 (2007).
83. Seki, E., Brenner, D. A. & Karin, M. A liver full of JNK: signaling in regulation of cell function and disease pathogenesis, and clinical approaches. *Gastroenterology* **143**, 307–320 (2012).
84. Jiang, H., Grenley, M. O., Bravo, M. J., Blumhagen, R. Z. & Edgar, B. A. EGFR/Ras/MAPK signaling mediates adult midgut epithelial homeostasis and regeneration in *Drosophila*. *Cell Stem Cell* **8**, 84–95 (2011).
85. Frey, M. R., Golovin, A. & Polk, D. B. Epidermal growth factor-stimulated intestinal epithelial cell migration requires Src family kinase-dependent p38 MAPK signaling. *J. Biol. Chem.* **279**, 44513–44521 (2004).
86. Ben-Neriah, Y. & Karin, M. Inflammation meets cancer, with NF- κ B as the matchmaker. *Nature Immunol.* **12**, 715–723 (2011).
87. Maeda, S. *et al.* IKK β is required for prevention of apoptosis mediated by cell-bound but not by circulating TNF α . *Immunity* **19**, 725–737 (2003).
88. Maeda, S., Kamata, H., Luo, J. L., Leffert, H. & Karin, M. IKK β couples hepatocyte death to cytokine-driven compensatory proliferation that promotes chemical hepatocarcinogenesis. *Cell* **121**, 977–990 (2005).
89. Greten, F. R. *et al.* IKK β links inflammation and tumorigenesis in a mouse model of colitis-associated cancer. *Cell* **118**, 285–296 (2004).
This is the first account of the crucial tumour-promoting function of NF- κ B signalling in intestinal epithelial cells and macrophages.
90. Egan, L. J. *et al.* I κ B-kinase β -dependent NF- κ B activation provides radioprotection to the intestinal epithelium. *Proc. Natl Acad. Sci. USA* **101**, 2452–2457 (2004).
91. Nenci, A. *et al.* Epithelial NEMO links innate immunity to chronic intestinal inflammation. *Nature* **446**, 557–561 (2007).
92. Eckmann, L. *et al.* Opposing functions of IKK β during acute and chronic intestinal inflammation. *Proc. Natl Acad. Sci. USA* **105**, 15058–15063 (2008).
93. Schwitala, S. *et al.* Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-like properties. *Cell* **152**, 25–38 (2013).
94. Jiang, H. & Edgar, B. A. Intestinal stem cell function in *Drosophila* and mice. *Curr. Opin. Genet. Dev.* **22**, 354–360 (2012).
95. Ernst, M., Thieme, S., Nguyen, P. M., Eissmann, M. & Putoczki, T. L. Epithelial gp130/Stat3 functions: an intestinal signaling node in health and disease. *Semin. Immunol.* **26**, 29–37 (2014).
96. Kolls, J. K., McCray, P. B. Jr & Chan, Y. R. Cytokine-mediated regulation of antimicrobial proteins. *Nature Rev. Immunol.* **8**, 829–835 (2008).
97. Wittkopf, N. *et al.* Activation of intestinal epithelial Stat3 orchestrates tissue defense during gastrointestinal infection. *PLoS ONE* **10**, e0118401 (2015).
98. Grivennikov, S. I. & Karin, M. Dangerous liaisons: STAT3 and NF- κ B collaboration and crosstalk in cancer. *Cytokine Growth Factor Rev.* **21**, 11–19 (2010).
99. Bollrath, J. *et al.* gp130-mediated STAT3 activation in enterocytes regulates cell survival and cell-cycle progression during colitis-associated tumorigenesis. *Cancer Cell* **15**, 91–102 (2009).
100. Anderson, C. A. *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nature Genet.* **43**, 246–252 (2011).
101. Moh, A. *et al.* Role of STAT3 in liver regeneration: survival, DNA synthesis, inflammatory reaction and liver mass recovery. *Lab. Invest.* **87**, 1018–1028 (2007).
102. He, G. *et al.* Hepatocyte IKK β /NF- κ B inhibits tumor promotion and progression by preventing oxidative stress-driven STAT3 activation. *Cancer Cell* **17**, 286–297 (2010).
103. Johnson, R. & Halder, G. The two faces of Hippo: targeting the Hippo pathway for regenerative medicine and cancer treatment. *Nature Rev. Drug Discov.* **13**, 63–79 (2014).
104. Baddour, L. M., Cha, Y. M. & Wilson, W. R. Clinical practice. Infections of cardiovascular implantable electronic devices. *N. Engl. J. Med.* **367**, 842–849 (2012).
105. Zhang, J. *et al.* YAP-dependent induction of amphiregulin identifies a non-cell-autonomous component of the Hippo pathway. *Nature Cell Biol.* **11**, 1444–1450 (2009).
106. Tschaharganeh, D. F. *et al.* Yes-associated protein up-regulates Jagged-1 and activates the Notch pathway in human hepatocellular carcinoma. *Gastroenterology* **144**, 1530–1542 (2013).
107. Yu, F. X. & Guan, K. L. The Hippo pathway: regulators and regulations. *Genes Dev.* **27**, 355–371 (2013).
108. Karpowicz, P., Perez, J. & Perrimon, N. The Hippo tumor suppressor pathway regulates intestinal stem cell regeneration. *Development* **137**, 4135–4145 (2010).
109. Ren, F. *et al.* Hippo signaling regulates *Drosophila* intestine stem cell

- proliferation through multiple pathways. *Proc. Natl Acad. Sci. USA* **107**, 21064–21069 (2010).
110. Shaw, R. L. *et al.* The Hippo pathway regulates intestinal stem cell proliferation during *Drosophila* adult midgut regeneration. *Development* **137**, 4147–4158 (2010).
 111. Staley, B. K. & Irvine, K. D. Warts and Yorkie mediate intestinal regeneration by influencing stem cell proliferation. *Curr. Biol.* **20**, 1580–1587 (2010).
 112. Zhou, D. *et al.* Mst1 and Mst2 protein kinases restrain intestinal stem cell proliferation and colonic tumorigenesis by inhibition of Yes-associated protein (Yap) overabundance. *Proc. Natl Acad. Sci. USA* **108**, e1312–e1320 (2011).
 113. Camargo, F. D. *et al.* YAP1 increases organ size and expands undifferentiated progenitor cells. *Curr. Biol.* **17**, 2054–2060 (2007).
 114. Cai, J. *et al.* The Hippo signaling pathway restricts the oncogenic potential of an intestinal regeneration program. *Genes Dev.* **24**, 2383–2388 (2010).
 115. Fre, S. *et al.* Notch signals control the fate of immature progenitor cells in the intestine. *Nature* **435**, 964–968 (2005).
 116. Stanger, B. Z., Datar, R., Murtaugh, L. C. & Melton, D. A. Direct regulation of intestinal fate by Notch. *Proc. Natl Acad. Sci. USA* **102**, 12443–12448 (2005).
 117. van Es, J. H. *et al.* Notch/ γ -secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature* **435**, 959–963 (2005).
- References 115 to 117 describe the crucial role of Notch signalling in the control of stem-cell fate in the mammalian gut.**
118. Micchelli, C. A. & Perrimon, N. Evidence that stem cells reside in the adult *Drosophila* midgut epithelium. *Nature* **439**, 475–479 (2006).
 119. Cordero, J. B. *et al.* c-Src drives intestinal regeneration and transformation. *EMBO J.* **33**, 1474–1491 (2014).
 120. Korinek, V. *et al.* Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nature Genet.* **19**, 379–383 (1998).
 121. Liu, S. *et al.* *Lgr4* gene deficiency increases susceptibility and severity of dextran sodium sulfate-induced inflammatory bowel disease in mice. *J. Biol. Chem.* **288**, 8794–8803 (2013).
 122. Ashton, G. H. *et al.* Focal adhesion kinase is required for intestinal regeneration and tumorigenesis downstream of Wnt/c-Myc signaling. *Dev. Cell* **19**, 259–269 (2010).
 123. Miyoshi, H., Ajima, R., Luo, C. T., Yamaguchi, T. P. & Stappenbeck, T. S. Wnt5a potentiates TGF- β signaling to promote colonic crypt regeneration after tissue injury. *Science* **338**, 108–113 (2012).
- This paper is an important account of the key parts played by Wnt and TGF β in control of intestinal regeneration.**
124. Clevers, H. C. & Bevins, C. L. Paneth cells: maestros of the small intestinal crypts. *Annu. Rev. Physiol.* **75**, 289–311 (2013).
 125. Potten, C. S. Extreme sensitivity of some intestinal crypt cells to X and γ irradiation. *Nature* **269**, 518–521 (1977).
 126. Sangiorgi, E. & Capecchi, M. R. Bmi1 is expressed *in vivo* in intestinal stem cells. *Nature Genet.* **40**, 915–920 (2008).
 127. Munoz, J. *et al.* The *Lgr5* intestinal stem cell signature: robust expression of proposed quiescent ‘+4’ cell markers. *EMBO J.* **31**, 3079–3091 (2012).
 128. Roche, K. C. *et al.* SOX9 maintains reserve stem cells and preserves radio-resistance in mouse small intestine. *Gastroenterology* **149**, 1553–1563 (2015).
 129. Buccacki, S. J. *et al.* Intestinal label-retaining cells are secretory precursors expressing *Lgr5*. *Nature* **495**, 65–69 (2013).
 130. Choi, T. Y., Ninov, N., Stainier, D. Y. & Shin, D. Extensive conversion of hepatic biliary epithelial cells to hepatocytes after near total loss of hepatocytes in zebrafish. *Gastroenterology* **146**, 776–788 (2014).
 131. Grompe, M. Liver stem cells, where art thou? *Cell Stem Cell* **15**, 257–258 (2014).
 132. Font-Burgada, J. *et al.* Hybrid periportal hepatocytes regenerate the injured liver without giving rise to cancer. *Cell* **162**, 766–779 (2015).
- This paper shows that periportal hepatocytes rather than oval cells are responsible for liver regeneration after injury, but do not give rise to cancer.**
133. Apte, U. *et al.* Wnt/ β -catenin signaling mediates oval cell response in rodents. *Hepatology* **47**, 288–295 (2008).
 134. Itoh, T., Kamiya, Y., Okabe, M., Tanaka, M. & Miyajima, A. Inducible expression of Wnt genes during adult hepatic stem/progenitor cell response. *FEBS Lett.* **583**, 777–781 (2009).
 135. Yang, W. *et al.* Wnt/ β -catenin signaling contributes to activation of normal and tumorigenic liver progenitor cells. *Cancer Res.* **68**, 4287–4295 (2008).
 136. Huch, M. *et al.* *In vitro* expansion of single *Lgr5*⁺ liver stem cells induced by Wnt-driven regeneration. *Nature* **494**, 247–250 (2013).
 137. Huch, M. *et al.* Long-term culture of genome-stable bipotent stem cells from adult human liver. *Cell* **160**, 299–312 (2015).
 138. Boulter, L. *et al.* Macrophage-derived Wnt opposes Notch signaling to specify hepatic progenitor cell fate in chronic liver disease. *Nature Med.* **18**, 572–579 (2012).
 139. Park, H. W. *et al.* Alternative Wnt signaling activates YAP/TAZ. *Cell* **162**, 780–794 (2015).
 140. Chung, A. S. *et al.* An interleukin-17-mediated paracrine network promotes tumor resistance to anti-angiogenic therapy. *Nature Med.* **19**, 1114–1123 (2013).

Acknowledgements The authors thank H. Gehart for producing the figures. Owing to space limitations, primary findings have been cited through reviews. Work by M.K., who is an American Cancer Society Research Professor and holder of the Ben and Wanda Hildyard Chair for Mitochondrial and Metabolic Diseases, is supported by the US National Institutes of Health, the Alliance for Lupus Research, the Lymphoma and Leukemia Society and the Superfund Basic Research Program. H.C. is supported by Stand Up to Cancer, the European Research Council, Alpe d’HuZes/KWF and the Netherlands Research Council NWO.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this paper at go.nature.com/dvpbdq. Correspondence should be addressed to M.K. (karinoffice@ucsd.edu) and H.C. (h.clevers@hubrecht.eu).

Angiocrine functions of organ-specific endothelial cells

Shahin Rafii¹, Jason M. Butler¹ & Bi-Sen Ding¹

Endothelial cells that line capillaries are not just passive conduits for delivering blood. Tissue-specific endothelium establishes specialized vascular niches that deploy sets of growth factors, known as angiocrine factors. These cues participate actively in the induction, specification, patterning and guidance of organ regeneration, as well as in the maintenance of homeostasis and metabolism. When upregulated following injury, they orchestrate self-renewal and differentiation of tissue-specific resident stem and progenitor cells into functional organs. Uncovering the mechanisms by which organotypic endothelium distributes physiological levels of angiocrine factors both spatially and temporally will lay the foundation for clinical trials that promote organ repair without scarring.

The microvascular circulation comprises a vast network of capillary endothelial cells (ECs) that connects the arteries to veins. These vascular beds, which are distinct from lymphatic vessels, were perceived as passive conduits with a responsibility for delivering oxygen and nutrients, modulating the coagulation of blood, regulating the transportation of inflammatory cells and serving as gatekeepers of cellular metabolism^{1,2}. However, these cells also perform other necessary physiological tasks: sustaining the homeostasis of resident stem cells and guiding the regeneration and repair of adult organs without provoking fibrosis.

These newly identified roles emerged from microanatomical observations, which revealed that epithelial, haematopoietic, mesenchymal and neuronal cells — together with their corresponding repopulating stem and progenitor cells — reside in proximity to capillary ECs. Genetic and biochemical studies have shown that ECs serve as a fertile, instructive niche that has important roles in homeostasis, metabolism and directing organ regeneration in a perfusion-independent manner. Tissue-specific ECs regulate these complex tasks by supplying the repopulating cells with stimulatory and inhibitory growth factors, morphogens, extracellular matrix components and chemokines. These EC-derived paracrine factors are defined collectively as angiocrine factors^{3,4} (Box 1).

The tissue-specific instructive functions of ECs have been demonstrated in studies showing that the deletion of angiocrine factors in adult ECs disrupts stem-cell homeostasis and impairs organ repair without compromising blood supply. Notably, intravenous transplantation and engraftment of tissue-specific ECs following injury augment organ reconstitution and function without instigating maladaptive fibrosis. On the basis of these observations, organotypic capillary ECs are now recognized as specialized niche cells that, through balanced physiological expression of angiocrine factors, maintain stem cells' capacity for quiescence and self-renewal. Spatially and temporally coordinated production of angiocrine factors after organ injury initiates and completes organ regeneration. This transformative model has opened a fresh chapter in translational vascular medicine. It has also raised the possibility that the inherent pro-regenerative potential of tissue-specific endothelium could be used therapeutically to orchestrate fibrosis-free healing and to restore homeostasis in tissues.

Although the angiocrine signals that guide the formation of the liver⁵ and pancreas⁶ in the fetus have been defined, the contribution of angiocrine signalling to the modulation of homeostasis and regeneration in adult organs has not been well studied until now. In this Review, we describe the instructive and inductive contributions of adult

tissue-specific ECs to the homeostatic and regenerative functions of repopulating stem and progenitor cells.

Instructive interactions of capillary ECs

The adult human body contains 10 trillion–60 trillion ECs that cover a vast surface area⁷. Tightly intertwined monolayers of ECs form the lumen of the blood circulatory system, which consists of large arteries, veins and extensively branched capillaries. Lymphatic vessels run parallel to capillaries and exist as an independent and open circulatory system. The surface area of capillaries represents more than 95% of the total circulatory surface area and arborizes into almost every cellular component of organs.

Tissue-specific stem and progenitor cells are strategically positioned in close proximity to homotypic capillary ECs (Fig. 1a–d). This intimate cellular interaction facilitates the delivery of membrane-bound and soluble angiocrine factors from specialized ECs to the recipient cells, which are located on the basolateral surface of blood vessels. Moreover, the luminal surface of ECs can serve as a signalling platform for stem and immune cells that navigate through the circulation. Tissue-resident parenchymal and stem cells regulate the activation state and response of ECs to regenerative stimuli through the production of angiogenic factors such as vascular endothelial growth factor (VEGF)-A, fibroblast growth factor (FGF)-2, stromal-cell-derived factor (SDF-1; also known as CXCL12), angiopoietins and thrombospondin-1 (TSP-1) (Fig. 1e). Thus, the capillary network — without the influence of pericytes and mesenchymal cells — provides an adaptive platform that has the functional plasticity to integrate and relay these intravascular and extravascular cues to both resting and regenerating organs.

Angiocrine-mediated self-renewal and differentiation

The formation of new blood vessels through angiogenesis is crucial to meet the metabolic demands of organs^{1,2}. Accumulating evidence indicates that ECs regulate organ homeostasis and repair through the production of angiocrine factors in an angiogenesis-independent manner (Box 1). The Greek philosopher and scientist Aristotle, who is widely considered to be the founder of classical biology, proposed that blood vessels direct the configuration of organs⁸. On pathophysiological stress (exposure to ionizing radiation, chemical injury or hypoxic conditions, for example) or loss of tissue mass, defined angiocrine factors emanate from activated ECs (Table 1). The activated ECs relay inflammatory and

¹Ansary Stem Cell Institute, Department of Medicine, Division of Regenerative Medicine, Weill Cornell Medical College, 1300 York Avenue, New York, New York 10065, USA.

injury-induced angiocrine signals to quiescent tissue-specific stem cells, which drives regeneration and enforces developmental set points to re-establish homeostatic conditions. Microvascular ECs therefore fulfil the criteria for professional niche cells that choreograph tissue regeneration by cradling and nurturing stem cells with physiological levels and proper stoichiometry of angiocrine factors. The contribution of the endothelial niche to mediating stem-cell homeostasis and function has been studied in depth in neural stem cells (NSCs), spermatogonial stem cells and haematopoietic stem and progenitor cells (HSPCs).

Neural stem cells

The adult brain contains two regions in which NSCs undergo neurogenesis: the ventricular subventricular zone (V-SVZ) and the subgranular zone (SGZ). In the V-SVZ, type B1 quiescent and activated NSCs give rise to type C transit amplifying cells and type A mature neuroblastic cell progenies, which are positioned in the proximity of capillary ECs^{9–12} (Fig. 2a). Similarly, in the SGZ, which is located in the dentate gyrus of the hippocampus, NSCs and their progenies reside near capillaries¹³. Brain capillaries are lined with ECs that are positive for VEGF receptor (VEGFR)-2 and vascular endothelial (VE)-cadherin, positive or negative for the CD133 antigen, negative for or express only low levels of thrombomodulin, and that display zones of variable permeability^{4,10,12} (Figs 1d, 2a). Subsets of V-SVZ and SGZ blood vessels have a specialized planar morphology in which NSCs extend their endfeet to contact ECs. This close proximity supports the possibility that angiocrine factors regulate neurogenesis.

In vitro studies of neuronal cells that were co-cultured with heterotypic-derived ECs support a model in which ECs regulate NSC homeostasis and differentiation. Primary human umbilical vein ECs have been shown to produce brain-derived nerve growth factor (BDNF), which fosters the expansion of neuroblasts¹⁴. Bovine pulmonary artery ECs and polyoma-mid-T-immortalized mouse brain capillary ECs, but not smooth-muscle cells, trigger Notch signalling by secreting soluble factors that increase the self-renewal of NSCs and drive neurogenesis⁹. Follow-up studies showed that pigment epithelium-derived factor (PEDF) was one of the secreted angiocrine factors that stimulates Notch-dependent self-renewing symmetric divisions of NSCs¹⁵.

Subsequent *in vivo* experiments demonstrated that angiocrine factors derived from brain ECs regulate the homeostasis and regeneration of NSCs both through direct cellular contact and in a paracrine manner^{13,16–18}. Under steady-state conditions, angiocrine expression of the membrane-bound proteins EphrinB2 and Jagged-1 (refs 19, 20) sustains the dormancy of quiescent NSCs. Direct contact of EC-derived EphrinB2 and Jagged-1 with the endfeet of these cells suppresses their entry into the cell cycle and keeps them in an undifferentiated state. Moreover, neurotrophin-3 (NT-3), which is selectively produced by ECs in the brain and choroid plexus, maintains NSC quiescence, in part, through the induction of endothelial nitric-oxide synthase and the production of nitric oxide^{21,22}. Although NSCs could also supply endothelial nitric-oxide synthase, the angiocrine release of NT-3 in the V-SVZ and cerebrospinal fluid dictates nitric oxide production that sustains stem-cell quiescence. Conditional deletion of NT-3 in adult mouse brain ECs depletes NT-3 in both cerebrospinal fluid and the V-SVZ, which leads to an increase in dividing activated NSCs that express glial fibrillary acidic protein (GFAP) and accelerates the exhaustion of the NSC pool. Thus, angiocrine factors actively enforce the quiescence that is crucial for the long-term maintenance of the NSC population.

During regenerative processes, irrigation of the V-SVZ by soluble angiocrine factors such as BDNF¹⁴, PEDF²³, betacellulin²⁴ and placental growth factor-2 (PlGF-2)²⁵, and of the SGZ by VEGF-C^{26,27}, orchestrates proliferation and differentiation of both quiescent and activated NSCs into transit amplifying cells and neuroblasts. Notably, graded angiocrine deposition of SDF-1 (ref. 28) and BDNF²⁹ by blood vessels that run along the rostral migratory stream in the mouse brain guides the proliferation of transit amplifying cells and their migration to the olfactory bulb³⁰. Therefore, brain capillary ECs not only supply the V-SVZ and SGZ with

BOX 1

Physiology of angiocrine factors

The paracrine factors produced by ECs that maintain organ homeostasis, balance the self-renewal and differentiation of stem cells and orchestrate organ regeneration and tumour growth are known as angiocrine factors. The term 'angiocrine' was created to emphasize the biological significance of the instructive factors produced by the ECs that influence the homeostasis of healthy and malignant tissues³. Angiocrine factors comprise secreted and membrane-bound inhibitory and stimulatory growth factors, trophogens, chemokines, cytokines, extracellular matrix components, exosomes and other cellular products that are supplied by tissue-specific ECs to help regulate homeostatic and regenerative processes in a paracrine or juxtacrine manner. These factors also play a part in adaptive healing and fibrotic remodelling.

Subsets of angiocrine factors can act as morphogens to determine the shape, architecture, size and patterning of regenerating organs. The angiocrine profile of each tissue-specific bed of ECs is different and reflects the diversity of cell types found adjacent to ECs in organs (Fig. 1a–e). Although subsets of angiocrine factors are produced constitutively, some angiogenic factors can modulate the production of other tissue-specific angiocrine factors. For example, VEGF-A induces the expression of defined angiocrine factors through interaction with VEGFR-1 and VEGFR-2 (Fig. 1e). Similarly, FGF-2 (through the activation of FGFR-1) and the angiopoietins (through their interaction with the receptor Tie2) drive the expression of unique clusters of angiocrine factors. TSP-1 functions in a complex manner and can act as an inhibitory angiogenic factor as well as directly influence the differentiation of stem and progenitor cells. The molecular programmes that govern the production of context-dependent angiocrine factors from organ-specific ECs remain undefined.

region-specific regenerative and path-finding cues, but also secrete angiocrine factors into cerebrospinal fluid to potentially modulate neuronal homeostasis throughout the brain.

Crosstalk between neuronal cells and angiogenic ECs allows the endothelial niche to adapt to regenerative neurogenesis (Fig. 1e). During vascular sprouting, cross-activation of ECs by neuronal-derived angiogenic factors regulates the differential production of angiocrine factors (Fig. 2a). After hypoxic injury, upregulation of VEGF-A through the activation of VEGFR-2 enhances the production of nitric oxide, which induces BDNF in brain capillary ECs to drive the expansion and maturation of transit amplifying cells¹⁶. Growth differentiation factor (GDF)-11 also enhances neurogenesis by remodelling the blood vessels³¹. Thus, endothelial niche cells in the brain possess a remarkable angiocrine plasticity that can adapt to the physiological demands of NSCs to initiate, execute and finalize neurogenic programmes.

Spermatogonial stem cells

Undifferentiated type A spermatogonial stem cells from mice reside in the vicinity of interstitial capillaries within the seminiferous tubules of the testes³². After perturbation of the testicular microenvironment, transplanted donor-derived spermatogonial stem cells localize to zones that are enriched in capillaries. *In vitro* studies have shown that spermatogonial stem cells that express the G-protein-coupled receptor GPR125 can directly convert to multipotent progenitor cells. Incubation of such spermatogonial stem cells with vascular-like stromal cells that carry the CD34 antigen is essential for the conversion of spermatogonial stem cells to pluripotent stem cells^{33–35}, and indicates that angiocrine factors play

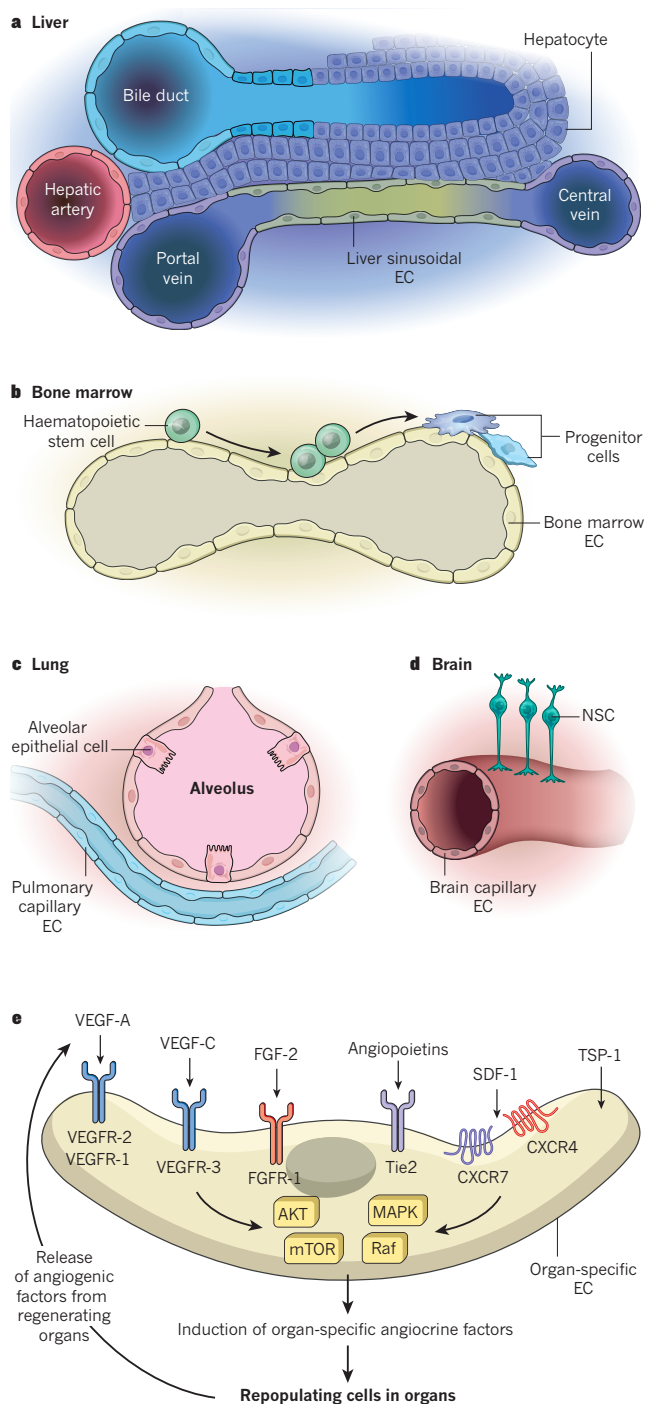


Figure 1 | Specialized structures of capillary ECs and their crosstalk with parenchymal and stem cells. **a**, In the liver, specialized fenestrated liver sinusoidal ECs are juxtaposed to hepatocytes. **b**, In haematopoietic organs, such as the bone marrow, arterial and sinusoidal EC-lined vessels are in direct contact with stem and progenitor cells. **c**, In the lungs, non-fenestrated pulmonary capillary ECs reside near alveolar epithelial cells. **d**, In the brain, capillaries composed of tightly connected ECs are found in the vicinity of NSCs. **e**, Under steady-state conditions or during angiogenesis, upregulation of angiogenic factors, such as VEGF-A through activation of VEGFR-2 and VEGFR-1, FGF-2 through activation of FGFR-1, and the angiopoietins through activation of the Tie2 receptor not only modulate angiogenic processes, but also trigger the expression of tissue-specific angiocrine factors. TSP-1 both tempers the angiogenic response and regulates the proliferation and differentiation of repopulating cells in organs, such as islet cells in the pancreas and alveolar epithelial cells in the lungs. AKT–mTOR and MAPK–Raf signalling pathways play a seminal part in inducing expression of organ-specific angiocrine factors.

an important part in regulating the maintenance and self-renewal of spermatogonial stem cells. Indeed, transcriptional analysis of testicular endothelium suggests that ECs could be a rich source of glial-cell-line-derived neurotrophic factor (GDNF)⁴. Further analysis of the phenotypic and functional properties of testicular ECs is necessary to determine the degree to which ECs influence spermatogonial stem-cell homeostasis by deploying angiocrine factors and depositing peritubular extracellular matrix components.

Haematopoietic stem cells

The first evidence that ECs establish an instructive niche for haematopoietic cells (Fig. 2b) was the demonstration that homotypic human bone-marrow-derived ECs expand human umbilical cord blood-derived CD34⁺ cells *ex vivo*^{36,37}. Furthermore, heterotypic primary ECs isolated from brain, heart and fetal tissues have since been shown to promote the proliferation of mouse^{38,39}, human^{40,41} and non-human primate HSPCs⁴². However, these co-culture studies were performed in media supplemented by serum that contained supraphysiological doses of growth factors and under ambient oxygen tension, which masked the full potential of ECs to regulate the function of the cells.

The development of techniques for serum-free and xenobiotic-free culture of primary human or mouse homotypic EC monolayers (Box 2) has facilitated the identification of angiocrine factors that support the self-renewal and differentiation of HSPCs in such co-culture studies^{43,44}. Co-culture studies have also been used to demonstrate that bone-marrow sinusoidal ECs that are positive for VEGFR-3, VEGFR-2, VE-cadherin and CD31 stimulate the self-renewal of HSPCs by expressing soluble and membrane-bound angiocrine factors^{45–47}, including bone morphogenic protein (BMP)2 and BMP4, insulin growth factor binding protein (IGFBP)2, SDF-1, Desert hedgehog (Dhh) protein, Notch ligands, Wingless-type MMTV integration site (Wnt)5a, and Kit ligand (Fig. 2b). Bone-marrow sinusoidal ECs also drive the lineage-specific differentiation of HSPCs by producing granulocyte macrophage colony-stimulating factor (GM-CSF), interleukin (IL)-6, IL-8, granulocyte colony-stimulating factor (G-CSF), IL-1, tumour necrosis factor (TNF), chemokines and metalloproteinases⁴⁵. Notably, ECs that are transitioning through various activation states also produce inhibitory factors, such as transforming growth factor (TGF)- β 1 (ref. 48), dickkopf-related protein (DKK)1 and DKK3, which block WNT signalling, and Noggin, which interferes with BMP signalling⁴⁵ (Fig. 2b, Table 1). Thus, ECs express inhibitory and stimulatory angiocrine factors that regulate the quiescence and proliferation of HSPCs.

ECs cultured under serum-free conditions were shown to supply angiocrine factors at physiological levels that increase the self-renewal of repopulating authentic mouse haematopoietic stem cells by 150-fold⁴⁶ and of human cord blood severe combined immunodeficiency repopulating cells by 8-fold⁴⁹. Direct contact between haematopoietic cells and ECs is essential for the self-renewal and differentiation of HSPCs^{45–47}. Compared with mesenchymal cells, ECs are more efficient at expanding umbilical cord blood-derived HSPCs⁵⁰. Other angiocrine factors, such as prostaglandin E2 (PGE2) (refs 51, 52), pleiotrophin⁵³ and epidermal growth factor (EGF)⁵⁴, drive haematopoietic reconstitution, which establishes ECs as a physiological repository of HSPC-supportive factors.

The first *in vivo* evidence to support the role of the endothelial niche in haematopoiesis came from a study of mice that are unable to produce soluble Kit ligand, an essential regulator of haematopoietic stem-cell biology⁵⁵. It demonstrated that compartmentalized — yet interactive — stromal and endothelial niche cells regulate the regeneration of HSPCs. In response to physiological stress, the activation of matrix metalloproteinase (MMP)-9 leads to the release of soluble Kit ligand from cells in the niche, which stimulates the regeneration and proper transportation of HSPCs. Follow-up studies showed that phenotypically marked stem cells reside in close proximity to the endothelial niche⁵⁶. Further evidence indicated that haematopoietic regeneration and thrombopoiesis after chemotherapy or irradiation is impaired by the conditional deletion of VEGFR-2 in ECs of adult mice⁴⁷ and by the targeting of VE-cadherin to disrupt reconstitution

Table 1 | Heterogeneity of angiocrine factors under steady-state conditions and during regeneration

Origin of EC	Repopulating cells	Regenerative model	Angiocrine factors	EC-specific gene knockout and emergent phenotype
Bone marrow	HSPCs	Steady state; chemotherapy, irradiation; angiogenesis	Kit ligand, SDF-1, IGFBP2, DKK1, DKK3, Wnt5a, Wnt2, TGF- β 1, Dhh, PGE2, BMP2, BMP4, Noggin AKT activation Kit ligand, Jagged-1, Dhh, IGFBP2, FGF-2, Ang2, EGF, Pleiotrophin AKT-MAPK activation IL-6, G-CSF, GM-CSF, M-CSF, Ang2, Jagged-1, Jagged-2, MMPs, IL-1, TNF	Kit ligand ⁶¹ , Sdf1 (ref. 62) Stem-cell depletion Jagged 1 (ref. 59) Impaired haematopoietic recovery and stem-cell depletion Vegfr2 (ref. 47) Angiocrine dysregulation, haematopoietic failure VE-cadherin (inhibition) ⁵⁷ Impaired haematopoietic recovery, thrombocytopaenia E-selectin ¹²⁴ Disruption to HSPC proliferation
Testicular capillary	GPR125 ⁺ germline stem cells	Chemically induced sterilization	GDNF, FGF-2	Not available
Brain capillary	NSCs, transit amplifying cells, neuroblasts	Steady state; regeneration; angiogenesis	Jagged-1, EphrinB2, NT-3, PEDF, betacellulin, VEGF-C, PIGF2, nitric oxide, BDNF	Jagged-1 (ref. 19), Ephrin B2 (ref. 19) Exhaustion of NSCs NT-3 (ref. 20) Decrease in NSC cycling
Liver sinusoid, central vein of the liver	Hepatocytes; stellate cells; Axin2 ⁺ Tbx3 ⁺ liver stem cells	70% partial hepatectomy; bile-duct ligation; acute and chronic injury through carbon tetrachloride; homeostatic conditions	Id1-Wnt2, Id1-HGF, Angiopoietin-2, Wnt2, Wnt9b, Rspodin3 Anti-fibrosis Apelin, Noggin, Follistatin-like 1 Pro-fibrosis TGF- β 1, BMP4	Angiopoietin 2 (ref. 85) Shift in recovery set point Id1 (ref. 81) Impaired regeneration Cxcr4 (ref. 86) Decrease in fibrotic healing Cxcr7 (ref. 86) Increase in pro-fibrotic healing Wntless ⁸² Defective regeneration and self-renewal R-spondin 3 (ref. 83) Impaired hepatic zonation
Pulmonary capillary	Lung alveolar epithelial cells, AT2 cells, basal epithelial cells, epithelial progenitor cells	Left-lung pneumonectomy; bleomycin	MMP-14-mediated EGFR ligand release, BMP4-NFATc1-TSP-1	Mmp14 (ref. 90) Abrogated neo-alveogenesis but intact capillary vasculature Vegfr2 (ref. 88), Fgfr1 (ref. 88) Impaired alveolar regeneration Tsp1 (ref. 93) Impaired epithelial regeneration
Type H and type L vessels in osteogenesis	Osteogenic cells	Steady state; regeneration	Noggin, BMPs, Jagged-1	Noggin ⁷⁴ Impaired osteogenesis
Pancreatic capillary	Islet cells, Pdx1 ⁺ islet progenitor cells	Streptozocin; pluripotent stem-cell derivatives	BMP4, BMP2, TSP-1, EGFL7	Egfl7 (ref. 100) Delayed islet-cell differentiation
White-fat capillary	Adipose stem cells	Steady state; regeneration	IGF1, IGFBP, TNF, IL-1	Not available
Cardiac-myocyte capillary	Myocytes	Coronary artery ligation	Neuregulin-1, Endothelin-1, nitric oxide	Neuregulin 1 (ref. 106) Impaired myocyte regeneration, loss of cardiac protection
Striated-muscle capillary	Pax7 ⁺ Myf5 ⁺ muscle satellite cells	Steady state; regeneration; post myotoxin delivery	IGF, HGF, FGF, PDGF-BB	Not available

of the endothelial niche^{57,58}. Therefore, the endothelial niche is essential not only for sustaining the self-renewal of haematopoietic stem cells, but also for multi-lineage reconstitution (Fig. 2b, Table 1).

Haematopoietic regeneration is orchestrated by the differential production of angiocrine factors that are induced by signalling pathways activated within ECs⁴⁵ (Fig. 2b). After myeloablative stress, angiogenic factors such as VEGF-A, VEGF-C, FGF-2 and the angiopoietins upregulate other angiocrine factors, including Jagged-1, through activation of AKT (also known as protein kinase B). Conditional deletion of Jagged-1 in ECs impairs haematopoietic recovery⁵⁹, which suggests that Notch activation prevents the exhaustion of HSPCs. During the angiogenic phase of regeneration, AKT phosphorylation is accompanied by the activation of p42/p44 mitogen-activated protein kinase (MAPK). This triggers the secretion of G-CSF, macrophage colony-stimulating factor (M-CSF), GM-CSF and IL-6 to expand populations of myeloid, megakaryocytic and lymphoid progenitor cells and aid haematopoietic reconstitution⁴⁵ (Fig. 2b). In turn, maturing haematopoietic cells produce inhibitory angiogenic factors that prevent excessive sprouting of regenerating sinusoidal vessels. For example, mature megakaryocytes produce TSP-1, which decelerates angiogenesis and shuts off the production of activating angiocrine factors to restore homeostasis^{46,60}. Notably, AKT-activated bone marrow ECs, which emulate some of the functions of *in vivo* angiogenic ECs, expand long-term repopulating haematopoietic stem cells under serum-free culture conditions, whereas bone-marrow-derived stromal cells direct stem-cell attrition⁴⁴. Moreover, protection of the haematopoietic microenvironment through transplantation of AKT-activated bone marrow ECs, but not mesenchymal ones, accelerates haematopoietic recovery after lethal irradiation⁴⁴. Therefore, the equilibrium between AKT and MAPK activation regulates multi-lineage haematopoietic recovery.

The contribution of the endothelial niche to steady-state

haematopoiesis was unravelled by studies in which selective deletion in ECs of SDF-1, Kit ligand or Jagged-1 impaired the maintenance of HSPCs^{44,61–64}. Several studies have also scrutinized the relative contribution of bone marrow perivascular cells to the homeostasis of HSPCs^{65–67}. Because the functions and structural stability of endothelial and non-vascular cells is mutually dependent, the deletion of factors in one niche has the potential to perturb the constituents of the neighbouring one. Therefore, genetic manipulations within the intimately associated endothelial niche and accompanying perivascular cells could have off-target effects, which must be taken into consideration. Nonetheless, the findings of these *in vivo* and reductionist *in vitro* studies suggest that, irrespective of the localization of HSPCs, angiocrine factors that are presented by either arteriolar or sinusoidal endothelial niches have executive functions and serve as 'rheostats' that choreograph haematopoietic stem-cell self-renewal and differentiation during homeostasis and recovery after haematopoietic suppression. Furthermore, these studies demonstrate that some, but not all, heterotypic ECs can support HSPC expansion, which confirms that each organotypic vascular bed is endowed with unique angiocrine attributes that are suitable for stem-cell homeostasis and reconstitution^{4,44,46,47}.

During fetal development, inductive signals from ECs⁶⁸ specify the development of haemogenic ECs^{69,70}. Thus, an endothelial niche could induce the direct conversion of all ECs into haemogenic ones, which give rise to definitive haematopoietic stem cells. Notably, endothelial niche-derived angiocrine signals are essential for the direct conversion of adult ECs into haematopoietic cells. In this approach, adult ECs were transduced with the transcription factors FosB, Gfi1, Runx1 and Spi1 (collectively termed FGRS). However, FGRS-transduced ECs failed to convert to engraftable haematopoietic cells unless they were co-cultured in direct contact with ECs⁷¹. Moreover, co-culture of haematopoietic cells that were derived from mouse and non-human primate pluripotent stem

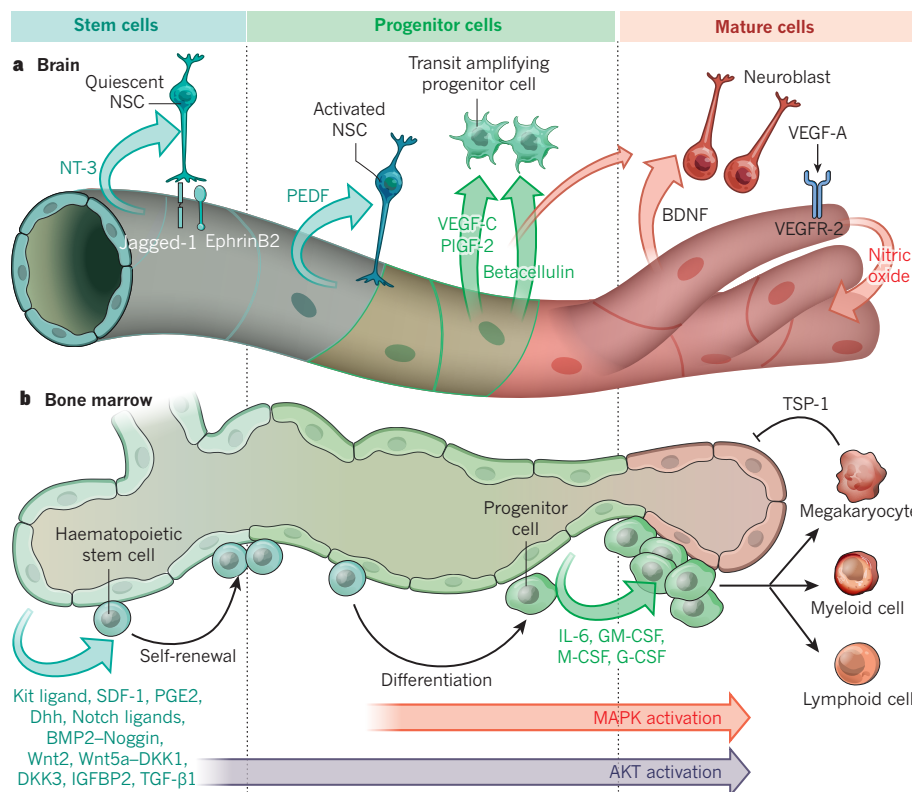


Figure 2 | Tissue-specific ECs orchestrate self-renewal and differentiation of stem and progenitor cells by supplying membrane-bound and secreted angiocrine factors. **a**, Within the V-SVZ and SGZ, brain capillary ECs manufacture specific angiocrine factors to promote the survival of quiescent NSCs (blue region). They also deploy PEDF, VEGF-C (mostly in the SGZ) and PIGF-2 to foster the proliferation of activated NSCs (green). Angiocrine secretion of betacellulin amplifies progenitor transit amplifying cells. During angiogenesis, activation of the VEGF-A–VEGFR-2 pathway induces the production of nitric oxide to upregulate BDNF, which acts with betacellulin to differentiate NSCs into neuroblasts and mature neurons that drive neurogenesis (red). **b**, Under steady-state conditions, a low level of AKT activation in bone marrow ECs stimulates the production of angiocrine factors that maintain and promote self-renewal of haematopoietic stem cells (blue). After myeloablative stress, inflicted by chemotherapy or irradiation, co-activation of AKT and MAPK initiates the expression of angiocrine factors that act on progenitor cells (green) to balance the differentiation of stem cells into lineage-committed progenitors such as myeloid and lymphoid cells. EC-derived Notch ligands prevent exhaustion of haematopoietic stem and progenitor cells. Expression of TSP-1 by maturing megakaryocytes halts ongoing haemangiogenesis to finalize the regeneration process (red).

cells and an endothelial niche enhanced the engraftment of putative haematopoietic cells, in part through the deployment of Notch ligands^{72,73}. Thus, angiocrine signals from ECs participate in the specification, development, homeostasis, self-renewal and differentiation of haematopoietic stem cells.

Angiogenesis and osteogenesis

The skeletal system is constantly being remodelled, yet the ratio of skeletal mass to haematopoietic volume remains constant. Although penetrating bone marrow arteries and sinusoidal vessels provide instructive signals for maintenance of the haematopoietic mass, specialized type H and type L bone capillaries modulate osteogenesis without compromising haematopoiesis^{74–76}. Type H vessels that are positive for CD31 and Endomucin regulate osteoblasts and the proliferation of chondrocytes^{74,75}. Type L vessels are an extension of type H vessels, and form sinusoidal vessels within the haematopoietic bone cavity. The angiocrine production of Noggin modulates skeletal patterning and ossification (Table 1). Notably, the expression of Noggin is downregulated by Notch signalling, which suggests that crosstalk between osteogenesis and haematopoiesis is conferred through the angiocrine expression of Jagged-1 (ref. 59). Hence, the balanced production of stimulatory and inhibitory angiocrine factors negotiates the allocation of space to bone and haematopoietic compartments.

Regeneration and fibrosis in the liver

Liver sinusoidal ECs distribute the intrahepatic blood flow between the hepatic artery, the hepatic vein and the portal vein⁷⁷. The tightly regulated hepatic blood flow requires these cells to perform as an adaptable plexus network. Notably, non-lymphatic liver sinusoidal ECs form fenestrated ECs that lack a basement membrane and are negative for CD34, positive for VEGFR-1, VEGFR-2, VEGFR-3, VE-cadherin and the chemokine receptor CXCR7, have low levels or a lack of CD31 and express factor VIII. They also express stabilin-2, CD32b, CD209L and lymphatic vessel endothelial hyaluronin acid receptor 1 (Lyve1) (refs 78–81). A combination of these markers has been used to isolate, characterize and cultivate non-lymphatic liver sinusoidal ECs⁸¹.

Under homeostatic conditions, angiocrine signals modulate the expansion of hepatocytes by enabling the proliferation of diploid Axin2- and

T-box transcription factor 3 (Tbx3)-positive cells that repopulate the liver⁸². These cells are located next to ECs in the central vein of the liver. Angiocrine production of Wnt2 and Wnt9b from these specialized ECs maintains Axin2 and Tbx3 double-positive hepatocytes that ultimately give rise to distal non-pericentral hepatocytes (Fig. 3a, Table 1). Deletion in ECs from adult mice of the gene *Wntless*, which encodes a specific transporter for Wnt-ligand secretion, depletes hepatic repopulating cells. Furthermore, specific angiocrine expression of the Wnt agonist Rspodin3 by ECs of the central vein of the liver — but not portal-vein ECs — establishes a β -catenin-dependent metabolic zonation⁸³. Therefore, the repopulating potential and metabolic zonation of hepatocytes are established by extrinsic angiocrine signals that are derived from an ‘inductive’ central-vein-of-liver endothelial niche.

The angiocrine contribution of liver sinusoidal ECs to liver regeneration can be studied through the surgical resection of up to 70% of the liver’s mass, known as a partial hepatectomy, which induces regeneration throughout the remaining liver. During the initial angiogenesis-independent inductive phase, which occurs 1–4 days after partial hepatectomy, VEGFR-2–AKT-dependent upregulation of transcription factor Id1 in non-proliferating liver sinusoidal ECs stimulates the expression of Wnt2 and hepatocyte growth factor (HGF) (Fig. 3a). Activation of VEGFR-1 on non-angiogenic liver sinusoidal ECs also induces the production of HGF and other pro-hepatic factors, such as heparin-binding EGF (HB-EGF), TGF- α and connective tissue growth factor (CTGF), to drive liver regeneration⁷⁹. On days 4–12 after partial hepatectomy, liver sinusoidal ECs promote proliferative angiogenesis to meet the metabolic demands of the enlarging liver (Fig. 3a). Putative bone-marrow-derived ECs that co-express CD133, CD45 and CD31 and have the capacity to produce HGF can also engraft into populations of regenerating liver sinusoidal ECs, which helps to boost regeneration of the liver⁸⁴.

Liver regeneration has a programmed set point to ensure restoration of the original hepatic mass that is modulated by liver sinusoidal ECs in an angiocrine-dependent manner⁸⁵. During the inductive phase of liver regeneration, Angiopoietin-2 is downregulated in liver sinusoidal ECs, which diminishes the expression of the inhibitory factor TGF- β 1. In the angiogenic phase of liver regeneration (4–8 days after partial hepatectomy), Angiopoietin-2 expression returns to normal levels to facilitate

VEGFR-2-dependent angiogenesis (Fig. 3a). Liver regeneration is delayed in adult mice in which Angiopoietin-2–Tie2 signalling is disrupted, which underscores the angiocrine function of Angiopoietin-2 in hepatic regeneration (Table 1).

Liver sinusoidal ECs are capable of either guiding the regeneration of the liver or engendering its pathological recovery by fibrosis, depending on differentially activated signalling pathways. After an acute insult to the liver, upregulation and activation of the SDF-1 receptor CXCR7 induces the angiocrine factors apelin and follistatin-like-1, which promote fibrosis-free repair⁸⁶ (Fig. 3a, Table 1). By contrast, chronic injury to the liver by repeated injection of carbon tetrachloride or bile-duct ligation leads to the upregulation of another SDF-1 receptor, CXCR4, and the suppression of CXCR7, which shifts the balance to angiocrine secretion of pro-fibrotic TGF- β 1 and BMP2, and leads to scarring. Thus, liver sinusoidal ECs are a cellular node that governs the regeneration, homeostasis and pathology of the liver.

Regeneration of the lung epithelium

The pulmonary circulation has an expansive capillary surface area that is covered by a thin layer of ECs. A delicate alveolar–capillary membrane is formed by the juxtaposition of these cells and alveolar epithelial cells to mediate gas exchange⁸⁷. This close association facilitates cellular crosstalk to modulate diverse pulmonary physiological processes. Using purification methods to eliminate lymphatic ECs (Box 2), pulmonary capillary ECs are identified by their expression of CD31, CD34, FGF receptor (FGFR)-1, VEGFR-1, and VEGFR-2, and their lack of CD45 (ref. 88).

The contribution of pulmonary capillary ECs to lung regeneration can be demonstrated by removal of the left lung in mammals. Known as a pneumonectomy, the procedure leads to a compensatory growth of lung mass that is driven by the expansion of alveolar epithelial stem and progenitor cells, which include alveolar type (AT)2 epithelial cells⁸⁹. After the pneumonectomy, the propagation of AT2 cells is elicited by MMP-14, which is induced in pulmonary capillary ECs⁸⁸. MMP-14 activates the EGF receptor (EGFR) on alveolar epithelial progenitor cells through its unmasking of the cryptic EGF-like motif in HB-EGF and the γ 2 chain of Laminin-5, and the proliferation of these cells achieves neo-alveogenesis (Fig. 3b, Table 1). Selective conditional deletion of FGFR-1 or VEGFR-2 from adult ECs impairs regeneration of the lung⁸⁸ (Table 1). Following the pneumonectomy, recruitment of platelets and delivery of SDF-1 to its receptors on pulmonary capillary ECs could also lead to the upregulation of MMP-14 in ECs⁹⁰. And selective conditional deletion of MMP-14 in adult ECs abrogates regeneration of the alveolar epithelium with negligible impact on vascular perfusion (Table 1). The progenitors of airway basal cells⁹¹ could also proliferate in response to angiocrine MMP-14 induction. FGF-2 and FGF-5 derived from cultured human basal cells stimulate the FGFR-1-dependent production of MMP-14 by pulmonary capillary ECs, which in turn supports the expansion and differentiation of basal cells⁹². Therefore, selective angiocrine upregulation of MMP-14 in pulmonary capillary ECs ignites the propagation of alveolar epithelial cells and basal cell progenitors.

Angiocrine signals also modulate the fate of other types of lung progenitor cell. BMP4 binds the BMP receptor BMPR-1 on ECs, which leads to NFATc1 transcriptional activation of TSP-1 and the differentiation of lung epithelial stem and progenitor cells⁹³ (Fig. 3b, Table 1). In 3D-homotypic co-cultures, angiocrine-derived TSP-1 stimulates the differentiation of lung epithelial progenitor cells, and in *Tsp1*-deficient mice, the differentiation of progenitor cells into AT1 and AT2 cells is impaired. Collectively, these studies demonstrate that the activation of VEGFR-2, FGFR-1 and BMPR-1 on pulmonary capillary ECs instructs neo-alveogenesis to restore gas-exchange function in the regenerating lungs, which sets the stage for the treatment of emphysema-like disorders with EC transplantation or angiocrine factors⁹⁴.

The modulation of metabolism

Molecular interactions between ECs and islet cells regulate pancreatic function. During the recuperation of islet cells from chemical injury,

BOX 2

In vitro endothelial niches

To propagate pure populations of organ-specific capillary ECs and to avoid their contamination with lymphatic ECs, monoclonal antibodies to EC-specific surface markers are infused intravenously into mice. The animals are killed after 10 minutes to avoid leakage of antibodies into the lymphatic vessels. Next, the intravitaly labelled ECs are isolated through enzymatic digestion of tissues and subsequent flow sorting to ensure the removal of pericytes and smooth-muscle cells^{4,43}. Purified ECs are then plated with angiogenic factors in the absence of pituitary extracts and serum to avoid the loss of tissue-specific signatures and to block the transition from endothelial to mesenchymal cells. However, these conditions can support the expansion of angiocrine-competent ECs only in the short term. To remove this hurdle, purified ECs can be transduced with lentiviral vectors that express myristoylated AKT or the non-oncogenic adenoviral gene *E4ORF1* (ref. 43). *E4ORF1* sustains the survival of ECs through specific activation of the AKT–mTOR pathway. ECs that express *E4ORF1* are not transformed and behave like primary ECs that ultimately undergo replicative senescence. This approach enables their long-term cultivation, with high viability in serum-free and xenobiotic-free conditions, which establishes an *ex vivo* platform for endothelial niches that recapitulates organotypic niches⁴. Monolayers of *E4ORF1*⁺ ECs sustain their organotypic pro-stem-cell functions, making them superior to ECs that have been immortalized by simian virus 40 large T antigen or polyomavirus middle T antigen, which fail to expand stem cells⁴⁵. Co-culture studies using *E4ORF1*⁺ ECs have helped to identify angiocrine pathways that foster the expansion and proliferation of parenchymal and stem cells. So far, *E4ORF1*⁺ ECs have not demonstrated any malignant potential and could be adapted to comply with regulatory clinical guidelines. Thus, these engineered ECs provide an ideal platform for interrogating the angiocrine function of ECs in models of organ regeneration and for translating the pro-regenerative therapeutic potential of these cells to the clinical setting.

angiocrine signals drive regeneration of the pancreas and the resolution of diabetes^{95,96}. An angiocrine supply of BMP2 and BMP4 is crucial for islet-cell homeostasis⁹⁶. Notably, the deposition of Laminin- α 4 chain by ECs stimulates insulin production by islet cells⁹⁷. Co-culture of islet cells with pancreatic ECs improves the survival of transplanted islet cells⁹⁸. Angiocrine delivery of TSP-1, through the modulation of TGF- β 1, promotes islet-cell proliferation⁹⁹. Angiocrine signals from ECs also guide the stage-specific differentiation of pluripotent stem cells into islet cells. For example, EC-derived EGF-like 7 (EGFL7) specifies the fate and proliferation of pancreas and duodenum homeobox (Pdx)1-positive islet-cell progenitors¹⁰⁰. Therefore, identification of the phenotypic and functional attributes of islet-specific ECs could unravel the angiocrine heterogeneity within pancreatic tissues and lead to new therapeutic strategies for islet-cell production.

Angiogenesis and transdifferentiation of ECs to adipogenic cells modulate thermogenesis and the remodelling of adipose tissue^{2,101}. Remarkably, ECs can also adjust metabolism in an angiogenesis-independent manner by deploying certain angiocrine factors. Adipose tissue regulates its mass by monitoring the differentiation of adipocyte stem cells into mature adipocytes. For example, in peroxisome proliferator-activated receptor- γ reporter mice, adipose stem cells were shown to embrace the white-fat vasculature, but not the vasculature of other tissues¹⁰². During adipogenesis, angiocrine signals such as TNE, IL-6, insulin growth factor (IGF)1 and IGF1s modulate the expansion and differentiation of adipocyte stem cells. In turn, adipocytes produce angiogenic factors that fine-tune the induction of angiocrine factors. Deciphering the molecular

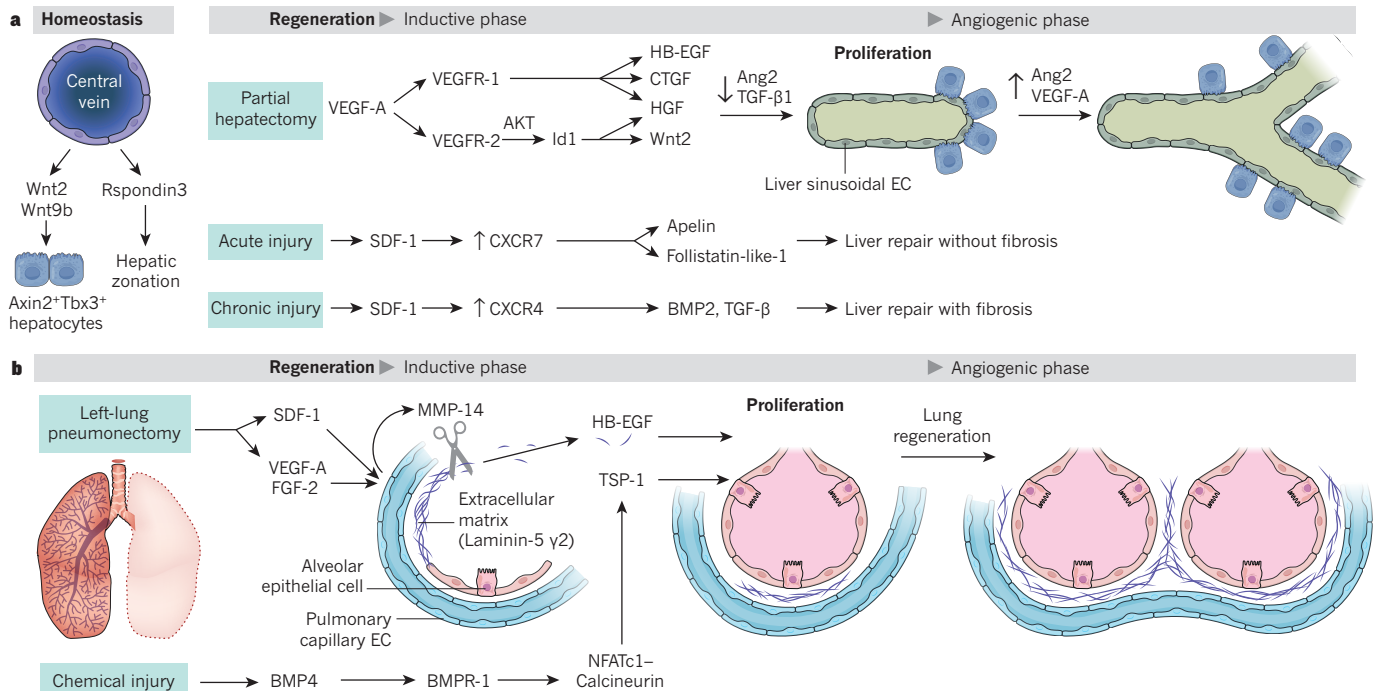


Figure 3 | Angiocrine signals support regeneration of hepatocytes and lung epithelial cells after injury. **a**, Wnt2 and Wnt9b produced by ECs in the central vein of the liver sustains the steady-state liver mass by replenishing the pool of Axin2⁺Tbx3⁺ hepatocytes. Rspodn3 produced by the same ECs establishes hepatic metabolic zonation. After partial hepatectomy, VEGF-A acts through VEGFR-2-mediated AKT activation to induce Id1, which upregulates HGF and Wnt2. It also upregulates HGF, HB-EGF and CTGF through VEGFR-1. Upregulation of these angiocrine factors, together with the downregulation of Angiopoietin-2 (Ang2) and TGF- β 1, unleashes angiogenesis-independent hepatocyte proliferation. Angiogenesis is stimulated on hepatic reconstitution that is finalized by the restoration of Ang2 and VEGF-A. After acute injury, activation of CXCR7 by SDF-1 triggers the expression of the anti-fibrotic factors apelin and follistatin-like-1. Chronic injury through persistent activation of

CXCR4 by SDF-1 and suppression of CXCR7 induces the pro-fibrotic factors BMP2 and TGF- β 1. The balance of CXCR4 and CXCR7 activation negotiates between liver regeneration and fibrosis. **b**, After left-lung pneumonectomy, VEGF-A, FGF-2 and SDF-1 switch on the expression of membrane-bound MMP-14. MMP-14 promotes the angiogenesis-independent regeneration of the alveolar epithelium by cleaving HB-EGF from the Laminin-5 γ 2-chain to unmask the cryptic EGF-receptor ligand. Recruitment of platelets to the regenerating lung delivers SDF-1 to pulmonary capillary ECs, which upregulates MMP-14 expression through the activation of CXCR4 (not shown). After chemical injury (with bleomycin, for example), BMP4 (through BMPR-1) activates the NFATc1–Calcineurin-dependent transcription of TSP-1 to facilitate epithelial proliferation. The increase in alveolar mass then triggers the angiogenic phase of lung regeneration.

details involved in the crosstalk between ECs and adipocyte stem cells has potential for the treatment of obesity, diabetes and metabolic syndromes.

Myogenic homeostasis

The idea that an endothelial niche modulates myogenesis stems from studies in which ECs were shown to facilitate the formation of striated muscle that was derived from pluripotent cells¹⁰³. Moreover, satellite striated-muscle precursor cells that express Pax7 and Myf5 are positioned in close proximity to ECs¹⁰⁴. Co-culture studies showed that ECs in heterotypic cultures support cycling of these satellite cells through the angiocrine production of IGF1, HGF, FGF-2, homodimers of platelet-derived growth factor B-chains (PDGF-BB) and VEGF-A (Table 1). Thus, homeostasis and regeneration of striated-muscle cells could also be coordinated by angiocrine signals.

Likewise, cardiac myocytes are surrounded by the unique capillary and specialized ECs that comprise the endocardium. Cardiac contractility and survival are controlled by EC-derived angiocrine factors, which include Neuregulin-1, endothelin-1 and nitric oxide¹⁰⁵. Neuregulin-1 is induced and released in response to hypoxic stress by the microvascular endocardial and subendocardial coronary arteries¹⁰⁶ (Table 1). Notably, deletion in adult ECs of the gene that encodes Neuregulin-1 decreases angiocrine-mediated protection of the ischaemic myocardium. Thus, after ischaemic injury, angiocrine factors execute regenerative programs that evoke cardiac repair and sustain myogenic-cell contractility.

Angiocrine factors as morphogens

Certain angiocrine factors, produced mainly by angiogenic ECs, consist of morphogens and chemokines that provide signals for proper tissue

patterning and direct repopulating cells to their predetermined anatomical destinations¹⁰⁷. During lung development, sprouting capillary ECs guide the patterning of alveolar epithelial cells¹⁰⁸. Similarly, after hepatotoxin-induced liver injury, the remaining liver sinusoidal ECs provide the cellular roadmap that steers the proper assembly of the hepatic architecture¹⁰⁹.

The size and activation state of the endothelial niche could also dictate the stem-cell mass and the extent to which an organ achieves its final volume and shape. Injury to ECs can decrease the surface area of the endothelial niche and deplete the stem-cell pool. This raises the possibility that an obligatory expansion in the mass of stem cells demands an equivalent increase in the surface area of the endothelial niche. Indeed, overexpression of VEGF-A in the V-SVZ of the brain increases the size of the endothelial niche, which augments neurogenesis^{17,110}. Whether the size of the endothelial niche in other organs is the main rheostat by which the mass of stem cells is set is an intriguing question that deserves further scrutiny.

Molecular determinants of angiocrine heterogeneity

In response to as yet unknown signals, organotypic ECs acquire unique structural features, such as fenestrations, and are induced to release specific sets of angiocrine factors^{4,111}. Notably, their response to various systemic endocrine factors, such as progesterone, is restricted to specific vascular boundaries — namely the uterine vascular bed¹¹². The mechanisms through which ECs acquire and maintain tissue-specific attributes could be regulated by extrinsic biophysical cues as well as by the expression of intrinsic cues that include certain transcription factors.

Specific groups of transcription factors, such as those that are encoded

by the homeobox, or *Hox*, genes, are expressed in organotypic ECs⁴. The differential expression of *Hox* genes provides positional identities to vascular zones at defined anatomic sites. In mice, *Hoxa3* and *Hoxc11*, which are associated with anterior and posterior domains, respectively, are differentially expressed in ECs along the body axis¹¹³. *Hox* gene products might also contribute to organotypic angiocrine expression. For example, *Hoxa9* triggers the expression of MMP-14 and EphrinB4, whereas *Hoxb5*, *Hoxb3* and *Hoxd3* control the positional expression of EphrinA1, VEGFR-2 and type I collagen¹¹⁴. It is plausible that extravascular cues, such as the biophysical forces that are imparted by shear stress and matrix elasticity, as well as crosstalk between parenchymal cells and ECs, enforce the expression of tissue-specific transcription factors in microvascular ECs and determine angiocrine heterogeneity.

Future directions

The emerging idea that organ homeostasis and regeneration are directed by tissue-specific microvascular ECs that function as instructive endothelial niches could be groundbreaking. This is because it focuses future work on what was conceived to be only a passive 'plumbing' system. These observations set the stage for capitalizing on the potential of specialized homotypic ECs as therapeutic drivers that can shepherd the regeneration of functional organs. However, the angiocrine profile of different organ-specific ECs is widely divergent, and a specific angiocrine profile must be established to elicit the desired effect on the local stem-cell population. Notably, the angiogenic and inflammatory states of ECs also play an important part in directing the proper production of angiocrine factors. Moreover, activation of signalling pathways other than AKT and MAPK — specifically, the Jak–Stat pathway — could switch on the expression of unique angiocrine factors. Therefore, to translate the potential of homotypic ECs to the clinical setting, the mechanism by which generic heterotypic ECs adopt tissue-specific angiocrine functions needs to be elucidated. The development of serum-free and xenobiotic-free co-culture platforms, such as E4ORF1⁺ ECs (Box 2), enables researchers to molecularly eavesdrop on the crosstalk between non-lymphatic homotypic or heterotypic ECs and tissue-repopulating stem cells. This could uncover the pathways that regulate the expression of tissue-specific angiocrine factors that are transcriptionally and biophysically induced.

In each organ, ECs perform the role of professional niche cells that elaborate a panoply of distinct sets of angiocrine factors that are distributed and delivered with precision to stem cells. Notably, ECs prevent stem cells from becoming exhausted or undergoing excessive self-renewal that could provoke the emergence of detrimental mutations. Our understanding of how stem cells integrate and interpret these myriad endothelial-derived signals to perform their organotypic functions is poor. Nonetheless, published data indicate that the coordinated presentation of a mixture of these angiocrine signals constitutes the long-sought holy grail of stem-cell self-renewal. Notably, the oscillatory release of inhibitory and stimulatory angiocrine factors might be crucial for the safe physiological self-renewal of engraftable stem cells, a process that is unlikely to be replicated by current approaches, in which stem cells are randomly pulsed with a cocktail of growth factors or small molecules. Thus, it is possible to translate the therapeutic potential of ECs for reconstituting pools of stem cells by recreating the appropriate endothelial niche, possibly through the transplantation of organotypic ECs.

The feasibility of this approach has been demonstrated by the intra-jugular transplantation of pulmonary capillary ECs into the regenerating alveoli of mice. By supplying the proper doses of angiocrine factors, these engrafted cells drive repopulation of epithelial cells and improve respiratory function^{88,90}. Infusion of human cord blood-derived ECs into neonatal mice with hyperoxia-induced alveolar injury also enhances the repair of alveoli¹¹⁵. The intravascular transplantation of liver sinusoidal ECs rescues fibrosis-free regeneration of the liver in Id1-deficient mice, which are refractory to liver regeneration⁸¹. In mice, transplantation of ECs also restores liver function after chemical injury^{77,84} and it corrects the phenotype of haemophilia A mice¹¹⁶. Administration of mouse brain capillary ECs^{117–119}, lung ECs¹¹⁹, bone marrow ECs⁴⁴ or other heterotypic ECs^{117–120}

enhances haematopoietic recovery after irradiation. Remarkably, senescent pancreatic islet cells regain their function¹²¹ when transplanted into young diabetic mice, which suggests that the infusion of young ECs could rejuvenate islet-cell function. Engineering clinical-grade tissue-specific or generic ECs^{43,122,123} that are capable of engrafting and adapting to an injured microenvironment, as well as producing the proper stoichiometry of angiocrine factors, will augment tissue regeneration and provide guidance for morphogenesis without instigating fibrosis.

It is plausible that the homeostasis and regenerative potential of almost every organ in mammals are regulated by angiocrine factors. Indeed, angiocrine signals might regulate the regenerative potential of stem cells in the hair-follicle bulge, Lgr5⁺ gastrointestinal repopulating cells, ovarian follicular cells, uterine syncytial cells, chondrocytes and tenocytes, as well as repopulating cells in endocrine organs and epithelial cells in the thymus and urogenital tissues. The development of technologies for isolating non-lymphatic capillary ECs (Box 2) from such sites will enable the angiocrine potential of ECs within these organs to be identified.

Although most of the studies described in this Review focused on angiocrine-derived growth factors and chemokines, the release of EC-derived exosomes and deposition of specific components of the extracellular matrix into the EC basal lamina, as well as the expression of adhesion molecules, such as E-selectin¹²⁴, could also modulate homeostatic and reparative processes. Two notable examples are the release of Laminin- α 4 by pancreatic ECs, which induces the production of insulin by islet cells¹²⁵, and MMP-14 generated by lung ECs through unmasking of the cryptic EGF-like motif in HB-EGF that is embedded within the Laminin-5 γ 2 chain⁸⁸. Similarly, extracellular matrix components that are laid down by organotypic ECs could also modulate tissue homeostasis by refining the affinity of adhesion molecules or by exposing growth factors that are embedded in the extracellular matrix.

Aberrant production of angiocrine factors could constitute the underlying pathogenesis of various conditions, such as cardiovascular or cerebrovascular diseases and the ageing process. Indeed, it is not known whether senescent ECs develop dysfunction in angiocrine production. An important topic not addressed in this Review is the mechanism by which inflammatory and immune cells could alter the normal angiocrine profile of ECs and thereby influence organ repair and scarring. The isolation and molecular profiling of ECs from the tissues of people with various disorders, such as individuals who are experiencing the complications of age-related and inflammatory diseases, could reveal the contribution of perturbed angiocrine signalling to these disorders. Restoration of the angiocrine function of maladapted or senescent vasculature, possibly through the transplantation of engineered ECs, could set the stage for treatments and help to diminish the economic burden of such disorders. ■

Received 9 August; accepted 26 November 2015.

1. Carmeliet, P. & Jain, R. K. Molecular mechanisms and clinical applications of angiogenesis. *Nature* **473**, 298–307 (2011).
 2. Ghesquière, B., Wong, B. W., Kuchnio, A. & Carmeliet, P. Metabolism of stromal and immune cells in health and disease. *Nature* **511**, 167–176 (2014).
 3. Butler, J. M., Kobayashi, H. & Rafii, S. Instructive role of the vascular niche in promoting tumour growth and tissue repair by angiocrine factors. *Nature Rev. Cancer* **10**, 138–146 (2010).
 4. Nolan, D. J. *et al.* Molecular signatures of tissue-specific microvascular endothelial cell heterogeneity in organ maintenance and regeneration. *Dev. Cell* **26**, 204–219 (2013).
- In this article, the authors developed an 'intravital' labelling approach to purify non-lymphatic mouse organ-specific ECs and used molecular profiling to demonstrate the remarkable angiocrine heterogeneity in ECs from various tissues.**
5. Matsumoto, K., Yoshitomi, H., Rossant, J. & Zaret, K. S. Liver organogenesis promoted by endothelial cells prior to vascular function. *Science* **294**, 559–563 (2001).
 6. Lammert, E., Cleaver, O. & Melton, D. Induction of pancreatic differentiation by signals from blood vessels. *Science* **294**, 564–567 (2001).
 7. Augustin, H. G., Kozian, D. H. & Johnson, R. C. Differentiation of endothelial cells: analysis of the constitutive and activated endothelial cell phenotypes. *Bioessays* **16**, 901–906 (1994).
 8. Crivellato, E., Nico, B. & Ribatti, D. Contribution of endothelial cells to organogenesis: a modern reappraisal of an old Aristotelian concept. *J. Anat.* **211**, 415–427 (2007).
 9. Shen, Q. *et al.* Endothelial cells stimulate self-renewal and expand neurogenesis

of neural stem cells. *Science* **304**, 1338–1340 (2004).

In this paper, the authors demonstrated that heterotypic ECs, but not stromal cells, support the expansion and differentiation of co-cultured NSCs by producing soluble angiocrine factors.

10. Shen, Q. *et al.* Adult SVZ stem cells lie in a vascular niche: a quantitative analysis of niche cell–cell interactions. *Cell Stem Cell* **3**, 289–300 (2008).
 11. Mirzadeh, Z., Merkle, F. T., Soriano-Navarro, M., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Neural stem cells confer unique pinwheel architecture to the ventricular surface in neurogenic regions of the adult brain. *Cell Stem Cell* **3**, 265–278 (2008).
 12. Tavazoie, M. *et al.* A specialized vascular niche for adult neural stem cells. *Cell Stem Cell* **3**, 279–288 (2008).
 13. Palmer, T. D., Willhoite, A. R. & Gage, F. H. Vascular niche for adult hippocampal neurogenesis. *J. Comp. Neurol.* **425**, 479–494 (2000).
 14. Leventhal, C., Rafii, S., Rafii, D., Shahar, A. & Goldman, S. A. Endothelial trophic support of neuronal production and recruitment from the adult mammalian subependyma. *Mol. Cell. Neurosci.* **13**, 450–464 (1999).
 15. Ramirez-Castillejo, C. *et al.* Pigment epithelium-derived factor is a niche signal for neural stem cell renewal. *Nature Neurosci.* **9**, 331–339 (2006).
 16. Goldman, S. A. & Chen, Z. Perivascular instruction of cell genesis and fate in the adult brain. *Nature Neurosci.* **14**, 1382–1389 (2011).
 17. Licht, T. & Keshet, E. The vascular niche in adult neurogenesis. *Mech. Dev.* **138**, 56–62 (2015).
 18. Silva-Vargas, V., Crouch, E. E. & Doetsch, F. Adult neural stem cells and their niche: a dynamic duo during homeostasis, regeneration, and aging. *Curr. Opin. Neurobiol.* **23**, 935–942 (2013).
 19. Ottone, C. *et al.* Direct cell–cell contact with the vascular niche maintains quiescent neural stem cells. *Nature Cell Biol.* **16**, 1045–1056 (2014).
 20. Ottone, C. & Parrinello, S. Multifaceted control of adult SVZ neurogenesis by the vascular niche. *Cell Cycle* **14**, 2222–2225 (2015).
 21. Delgado, A. C. *et al.* Endothelial NT-3 delivered by vasculature and CSF promotes quiescence of subependymal neural stem cells through nitric oxide induction. *Neuron* **83**, 572–585 (2014).
 22. Silva-Vargas, V. & Doetsch, F. A new twist for neurotrophins: endothelial-derived NT-3 mediates adult neural stem cell quiescence. *Neuron* **83**, 507–509 (2014).
 23. Andreu-Agulló, C., Morante-Redolat, J. M., Delgado, A. C. & Farinas, I. Vascular niche factor PEDF modulates Notch-dependent stemness in the adult subependymal zone. *Nature Neurosci.* **12**, 1514–1523 (2009).
 24. Gómez-Gaviró, M. V. *et al.* Betacellulin promotes cell proliferation in the neural stem cell niche and stimulates neurogenesis. *Proc. Natl Acad. Sci. USA* **109**, 1317–1322 (2012).
 25. Crouch, E. E., Liu, C., Silva-Vargas, V. & Doetsch, F. Regional and stage-specific effects of prospectively purified vascular cells on the adult V-SVZ neural stem cell lineage. *J. Neurosci.* **35**, 4528–4539 (2015).
 26. Han, J. *et al.* Vascular endothelial growth factor receptor 3 controls neural stem cell activation in mice and humans. *Cell Rep.* **10**, 1158–1172 (2015).
 27. Le Bras, B. *et al.* VEGF-C is a trophic factor for neural progenitors in the vertebrate embryonic brain. *Nature Neurosci.* **9**, 340–348 (2006).
 28. Kokovay, E. *et al.* Adult SVZ lineage cells home to and leave the vascular niche via differential responses to SDF1/CXCR4 signaling. *Cell Stem Cell* **7**, 163–173 (2010).
 29. Snayman, M. *et al.* Vasculature guides migrating neuronal precursors in the adult mammalian forebrain via brain-derived neurotrophic factor signaling. *J. Neurosci.* **29**, 4172–4188 (2009).
 30. Whitman, M. C., Fan, W., Relu, L., Rodríguez-Gil, D. J. & Greer, C. A. Blood vessels form a migratory scaffold in the rostral migratory stream. *J. Comp. Neurol.* **516**, 94–104 (2009).
 31. Katsimpardi, L. *et al.* Vascular and neurogenic rejuvenation of the aging mouse brain by young systemic factors. *Science* **344**, 630–634 (2014).
 32. Yoshida, S., Sukeno, M. & Nabeshima, Y. A vasculature-associated niche for undifferentiated spermatogonia in the mouse testis. *Science* **317**, 1722–1726 (2007).
 33. Seandel, M. *et al.* Generation of functional multipotent adult stem cells from GPR125⁺ germ-line progenitors. *Nature* **449**, 346–350 (2007).
 34. Seandel, M. *et al.* Niche players: spermatogonial progenitors marked by GPR125. *Cell Cycle* **7**, 135–140 (2008).
 35. Kim, J., Seandel, M., Falcioni, I., Wen, D. & Rafii, S. CD34⁺ testicular stromal cells support long-term expansion of embryonic and adult stem and progenitor cells. *Stem Cells* **26**, 2516–2522 (2008).
 36. Rafii, S. *et al.* Isolation and characterization of human bone marrow microvascular endothelial cells: hematopoietic progenitor cell adhesion. *Blood* **84**, 10–19 (1994).
 37. Rafii, S. *et al.* Human bone marrow microvascular endothelial cells support long-term proliferation and differentiation of myeloid and megakaryocytic progenitors. *Blood* **86**, 3353–3363 (1995).
- This was the first report of homotypic co-culture of human bone marrow ECs and human hematopoietic cells, and it underscored the capacity of endothelial niches to support the expansion of hematopoietic cells through the production of angiocrine factors in the absence of exogenous factors.**
38. Li, W. *et al.* Primary endothelial cells isolated from the yolk sac and para-aortic splanchnopleura support the expansion of adult marrow stem cells *in vitro*. *Blood* **102**, 4345–4353 (2003).
 39. Li, W., Johnson, S. A., Shelley, W. C. & Yoder, M. C. Hematopoietic stem cell repopulating ability can be maintained *in vitro* by some primary endothelial cells. *Exp. Hematol.* **32**, 1226–1237 (2004).
 40. Chute, J. P., Saini, A. A., Kampen, R. L., Wells, M. R. & Davis, T. A. A comparative study of the cell cycle status and primitive cell adhesion molecule profile of human CD34⁺ cells cultured in stroma-free versus porcine microvascular endothelial cell cultures. *Exp. Hematol.* **27**, 370–379 (1999).
 41. Chute, J. P. *et al.* *Ex vivo* culture with human brain endothelial cells increases the SCID-repopulating capacity of adult human bone marrow. *Blood* **100**, 4433–4439 (2002).
 42. Brandt, J. E. *et al.* *Ex vivo* expansion of autologous bone marrow CD34⁺ cells with porcine microvascular endothelial cells results in a graft capable of rescuing lethally irradiated baboons. *Blood* **94**, 106–113 (1999).
 43. Seandel, M. *et al.* Generation of a functional and durable vascular niche by the adenoviral *E4ORF1* gene. *Proc. Natl Acad. Sci. USA* **105**, 19288–19293 (2008).
- This report demonstrated that transduction of a non-oncogenic adenoviral *E4ORF1* gene enables the serum-free and xenobiotic-free expansion of primary human ECs, sustaining their angiocrine repertoire.**
44. Poulos, M. G. *et al.* Vascular platform to define hematopoietic stem cell factors and enhance regenerative hematopoiesis. *Stem Cell Reports* **10**, 881–894 (2015).
 45. Kobayashi, H. *et al.* Angiocrine factors from Akt-activated endothelial cells balance self-renewal and differentiation of haematopoietic stem cells. *Nature Cell Biol.* **12**, 1046–1056 (2010).
 46. Butler, J. M. *et al.* Endothelial cells are essential for the self-renewal and repopulation of Notch-dependent hematopoietic stem cells. *Cell Stem Cell* **6**, 251–264 (2010).
- This article employs a serum-free endothelial niche platform to demonstrate that ECs foster the *ex vivo* expansion and self-renewal of Notch-dependent, authentic long-term repopulating hematopoietic stem cells in mice.**
47. Hooper, A. T. *et al.* Engraftment and reconstitution of hematopoiesis is dependent on VEGFR2-mediated regeneration of sinusoidal endothelial cells. *Cell Stem Cell* **4**, 263–274 (2009).
 48. Brenet, F., Kermani, P., Spektor, R., Rafii, S. & Scandura, J. M. TGFβ restores hematopoietic homeostasis after myelosuppressive chemotherapy. *J. Exp. Med.* **210**, 623–639 (2013).
 49. Butler, J. M. *et al.* Development of a vascular niche platform for expansion of repopulating human cord blood stem and progenitor cells. *Blood* **120**, 1344–1347 (2012).
 50. Raynaud, C. M. *et al.* Endothelial cells provide a niche for placental hematopoietic stem/progenitor cell expansion through broad transcriptomic modification. *Stem Cell Res.* **11**, 1074–1090 (2013).
 51. North, T. E. *et al.* Prostaglandin E2 regulates vertebrate hematopoietic stem cell homeostasis. *Nature* **447**, 1007–1011 (2007).
 52. Hoggatt, J. *et al.* Differential stem- and progenitor-cell trafficking by prostaglandin E2. *Nature* **495**, 365–369 (2013).
 53. Himburg, H. A. *et al.* Pleiotrophin regulates the retention and self-renewal of hematopoietic stem cells in the bone marrow vascular niche. *Cell Reports* **2**, 964–975 (2012).
 54. Doan, P. L. *et al.* Epidermal growth factor regulates hematopoietic regeneration after radiation injury. *Nature Med.* **19**, 295–304 (2013).
 55. Heissig, B. *et al.* Recruitment of stem and progenitor cells from the bone marrow niche requires MMP-9 mediated release of Kit-ligand. *Cell* **109**, 625–637 (2002).
- This paper was the first to propose that endothelial and non-vascular niches cooperate by establishing dynamic intertwined niches that facilitate the reconstitution of haematopoiesis and stem-cell transportation through an MMP-9-mediated increase in the bioavailability of soluble Kit ligand.**
56. Kiel, M. J., Yilmaz, O. H., Iwashita, T., Terhorst, C. & Morrison, S. J. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109–1121 (2005).
 57. Avecilla, S. T. *et al.* Chemokine-mediated interaction of hematopoietic progenitors with the bone marrow vascular niche is required for thrombopoiesis. *Nature Med.* **10**, 64–71 (2004).
 58. Hamada, T. *et al.* Transendothelial migration of megakaryocytes in response to stromal cell-derived factor 1 (SDF-1) enhances platelet formation. *J. Exp. Med.* **188**, 539–548 (1998).
 59. Poulos, M. G. *et al.* Endothelial Jagged-1 is necessary for homeostatic and regenerative hematopoiesis. *Cell Reports* **4**, 1022–1034 (2013).
 60. Kopp, H. G. *et al.* Thrombospondins deployed by thrombopoietic cells determine angiogenic switch and extent of revascularization. *J. Clin. Invest.* **116**, 3277–3291 (2006).
 61. Ding, L., Saunders, T. L., Enikolopov, G. & Morrison, S. J. Endothelial and perivascular cells maintain haematopoietic stem cells. *Nature* **481**, 457–462 (2012).
- In this article, the authors showed that the angiocrine supply of Kit ligand by ECs is essential for hematopoietic maintenance because the selective conditional deletion of Kit ligand in ECs results in impaired haematopoiesis.**
62. Ding, L. & Morrison, S. J. Hematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches. *Nature* **495**, 231–235 (2013).
 63. Inra, C. N. *et al.* A perisinusoidal niche for extramedullary haematopoiesis in the spleen. *Nature* **527**, 466–471 (2015).
 64. Kimura, Y. *et al.* c-Kit-mediated functional positioning of stem cells to their niches is essential for maintenance and regeneration of adult hematopoiesis. *PLoS ONE* **6**, e26918 (2011).
 65. Kunisaki, Y. *et al.* Arteriolar niches maintain haematopoietic stem cell quiescence. *Nature* **502**, 637–643 (2013).
 66. Morrison, S. J. & Scadden, D. T. The bone marrow niche for hematopoietic stem cells. *Nature* **505**, 327–334 (2014).
 67. Acar, M. *et al.* Deep imaging of bone marrow shows non-dividing stem cells are mainly perisinusoidal. *Nature* **526**, 126–130 (2015).
 68. Nguyen, P. D. *et al.* Hematopoietic stem cell induction by somite-derived endothelial cells controlled by *meox1*. *Nature* **512**, 314–318 (2014).
 69. Medvinsky, A. & Dzierzak, E. Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell* **86**, 897–906 (1996).

70. Chen, M. J., Yokomizo, T., Zeigler, B. M., Dzierzak, E. & Speck, N. A. Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature* **457**, 887–891 (2009).
71. Sandler, V. M. *et al.* Reprogramming human endothelial cells to haematopoietic cells requires vascular induction. *Nature* **511**, 312–318 (2014).
72. Gori, J. L. *et al.* Vascular niche promotes hematopoietic multipotent progenitor formation from pluripotent stem cells. *J. Clin. Invest.* **125**, 1243–1254 (2015).
73. Hadland, B. K. *et al.* Endothelium and NOTCH specify and amplify aorta-gonad-mesonephros-derived hematopoietic stem cells. *J. Clin. Invest.* **125**, 2032–2045 (2015).
74. Ramasamy, S. K., Kusumbe, A. P., Wang, L. & Adams, R. H. Endothelial Notch activity promotes angiogenesis and osteogenesis in bone. *Nature* **507**, 376–380 (2014).
75. Ramasamy, S. K., Kusumbe, A. P. & Adams, R. H. Regulation of tissue morphogenesis by endothelial cell-derived signals. *Trends Cell Biol.* **25**, 148–157 (2015).
76. Kusumbe, A. P., Ramasamy, S. K. & Adams, R. H. Coupling of angiogenesis and osteogenesis by a specific vessel subtype in bone. *Nature* **507**, 323–328 (2014).
77. DeLeve, L. D. Liver sinusoidal endothelial cells and liver regeneration. *J. Clin. Invest.* **123**, 1861–1866 (2013).
78. Falkowski, M., Schledzewski, K., Hansen, B. & Goerdts, S. Expression of stabilin-2, a novel fasciclin-like hyaluronan receptor protein, in murine sinusoidal endothelia, avascular tissues, and at solid/liquid interfaces. *Histochem. Cell Biol.* **120**, 361–369 (2003).
79. LeCouter, J. *et al.* Angiogenesis-independent endothelial protection of liver: role of VEGFR-1. *Science* **299**, 890–893 (2003).
80. DeLeve, L. D. Liver sinusoidal endothelial cells in hepatic fibrosis. *Hepatology* **61**, 1740–1746 (2015).
81. Ding, B. S. *et al.* Inductive angiocrine signals from sinusoidal endothelium are required for liver regeneration. *Nature* **468**, 310–315 (2010).
82. Wang, B., Zhao, L., Fish, M., Logan, C. Y. & Nusse, R. Self-renewing diploid Axin2⁺ cells fuel homeostatic renewal of the liver. *Nature* **524**, 180–185 (2015).
- This paper shows that under steady-state conditions, angiocrine deployment of Wnt2 and Wnt9b by ECs of the central vein of the liver enables the generation of diploid repopulating hepatocytes that maintain liver mass.**
83. Rocha, A. S. *et al.* The angiocrine factor Rspn3 is a key determinant of liver zonation. *Cell Reports* **12**, 1757–1764 (2015).
84. Wang, L., Wang, X., Xie, G., Hill, C. K. & DeLeve, L. D. Liver sinusoidal endothelial cell progenitor cells promote liver regeneration in rats. *J. Clin. Invest.* **122**, 1567–1573, (2012).
85. Hu, J. *et al.* Endothelial cell-derived Angiopoietin-2 controls liver regeneration as a spatiotemporal rheostat. *Science* **343**, 416–419 (2014).
- This report provides genetic evidence to show that angiocrine-derived Angiopoietin-2 induces angiogenesis-independent hepatic proliferation and finalizes angiogenesis-dependent liver regeneration by activating its cognate Tie2 receptor and by regulating TGF- β .**
86. Ding, B. S. *et al.* Divergent angiocrine signals from vascular niche balance liver regeneration and fibrosis. *Nature* **505**, 97–102 (2014).
87. Morrisey, E. E. & Hogan, B. L. Preparing for the first breath: genetic and cellular mechanisms in lung development. *Dev. Cell* **18**, 8–23 (2010).
88. Ding, B. S. *et al.* Endothelial-derived angiocrine signals induce and sustain regenerative lung alveolarization. *Cell* **147**, 539–553 (2011).
89. Hogan, B. L. *et al.* Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function. *Cell Stem Cell* **15**, 123–138 (2014).
90. Rafii, S. *et al.* Platelet-derived SDF-1 primes the pulmonary capillary vascular niche to drive lung alveolar regeneration. *Nature Cell Biol.* **17**, 123–136 (2015).
91. Rock, J. R. *et al.* Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc. Natl Acad. Sci. USA* **106**, 12771–12775 (2009).
92. Ding, B. S., Gomi, K., Rafii, S., Crystal, R. G. & Walters, M. S. Endothelial MMP14 is required for endothelial dependent growth support of human airway basal cells. *J. Cell Sci.* **128**, 2983–2988 (2015).
93. Lee, J. H. *et al.* Lung stem cell differentiation in mice directed by endothelial cells via a BMP4–NFATc1–Thrombospondin-1 axis. *Cell* **156**, 440–455 (2014).
94. Petrache, I. *et al.* Ceramide upregulation causes pulmonary cell apoptosis and emphysema-like disease in mice. *Nature Med.* **11**, 491–498 (2005).
95. Rivas-Carrillo, S. D. *et al.* Endothelial cells promote pancreatic stem cell activation during islet regeneration in mice. *Transplant. Proc.* **43**, 3209–3211 (2011).
96. Talavera-Adame, D. & Dafoe, D. C. Endothelium-derived essential signals involved in pancreas organogenesis. *World J. Exp. Med.* **5**, 40–49 (2015).
97. Nikolova, G. *et al.* The vascular basement membrane: a niche for insulin gene expression and β cell proliferation. *Dev. Cell* **10**, 397–405 (2006).
98. Pan, X. *et al.* Islet graft survival and function: concomitant culture and transplantation with vascular endothelial cells in diabetic rats. *Transplantation* **92**, 1208–1214 (2011).
99. Olerud, J. *et al.* Thrombospondin-1: an islet endothelial cell signal of importance for β -cell function. *Diabetes* **60**, 1946–1954 (2011).
100. Kao, D. I. *et al.* Endothelial cells control pancreatic cell fate at defined stages through EGFL7 signaling. *Stem Cell Reports* **4**, 181–189 (2015).
101. Cao, Y. Angiogenesis and vascular functions in modulation of obesity, adipose metabolism, and insulin sensitivity. *Cell Metab.* **18**, 478–489 (2013).
102. Tang, W. *et al.* White fat progenitor cells reside in the adipose vasculature. *Science* **322**, 583–586 (2008).
103. Shmelkov, S. V. *et al.* Cytokine preconditioning promotes codifferentiation of human fetal liver CD133⁺ stem cells into angiomyogenic tissue. *Circulation* **111**, 1175–1183 (2005).
104. Christov, C. *et al.* Muscle satellite cells and endothelial cells: close neighbors and privileged partners. *Mol. Biol. Cell* **18**, 1397–1409 (2007).
105. Noireaud, J. & Andriantsitohaina, R. Recent insights in the paracrine modulation of cardiomyocyte contractility by cardiac endothelial cells. *Biomed. Res. Int.* **2014**, 923805 (2014).
106. Hedhli, N. *et al.* Endothelium-derived neuregulin protects the heart against ischemic injury. *Circulation* **123**, 2254–2262 (2011).
- In this paper, the authors provide genetic evidence to demonstrate that selective deletion of Neuregulin-1 in ECs from adult mice impairs myogenic repair, which implies that an angiocrine supply of Neuregulin-1 is essential for cardiac myocyte protection and regeneration.**
107. Cleaver, O. & Melton, D. A. Endothelial signaling during development. *Nature Med.* **9**, 661–668 (2003).
108. Lazarus, A. L. *et al.* A perfusion-independent role of blood vessels in determining branching stereotypy of lung airways. *Development* **138**, 2359–2368 (2011).
109. Hoehme, S. *et al.* Prediction and validation of cell alignment along microvessels as order principle to restore tissue architecture in liver regeneration. *Proc. Natl Acad. Sci. USA* **107**, 10371–10376 (2010).
110. Licht, T. *et al.* Reversible modulations of neuronal plasticity by VEGF. *Proc. Natl Acad. Sci. USA* **108**, 5081–5086 (2011).
111. Aird, W. C. Endothelial cell heterogeneity. *Cold Spring Harb. Perspect. Med.* **2**, a006429 (2012).
112. Goddard, L. M. *et al.* Progesterone receptor in the vascular endothelium triggers physiological uterine permeability preimplantation. *Cell* **156**, 549–562 (2014).
113. Pruett, N. D. *et al.* Evidence for Hox-specified positional identities in adult vasculature. *BMC Dev. Biol.* **8**, 93 (2008).
114. Douville, J. M. & Wigle, J. T. Regulation and function of homeodomain proteins in the embryonic and adult vascular systems. *Can. J. Physiol. Pharmacol.* **85**, 55–65 (2007).
115. Alphonse, R. S. *et al.* Existence, functional impairment, and lung repair potential of endothelial colony-forming cells in oxygen-induced arrested alveolar growth. *Circulation* **129**, 2144–2157 (2014).
116. Follenzi, A. *et al.* Transplanted endothelial cells repopulate the liver endothelium and correct the phenotype of hemophilia A mice. *J. Clin. Invest.* **118**, 935–945 (2008).
117. Salter, A. B. *et al.* Endothelial progenitor cell infusion induces hematopoietic stem cell reconstitution in vivo. *Blood* **113**, 2104–2107 (2009).
118. Chute, J. P. *et al.* Transplantation of vascular endothelial cells mediates the hematopoietic recovery and survival of lethally irradiated mice. *Blood* **109**, 2365–2372 (2007).
119. Li, B. *et al.* Endothelial cells mediate the regeneration of hematopoietic stem cells. *Stem Cell Res.* **4**, 17–24 (2010).
120. Zachman, D. K. *et al.* Endothelial cells mitigate DNA damage and promote the regeneration of hematopoietic stem cells after radiation injury. *Stem Cell Res.* **11**, 1013–1021 (2013).
121. Almacá, J. *et al.* Young capillary vessels rejuvenate aged pancreatic islets. *Proc. Natl Acad. Sci. USA* **111**, 17612–17617 (2014).
122. Ginsberg, M. *et al.* Efficient direct reprogramming of mature amniotic cells into endothelial cells by ETS factors and TGF β suppression. *Cell* **151**, 559–575 (2012).
123. James, D. *et al.* Expansion and maintenance of human embryonic stem cell-derived endothelial cells by TGF β inhibition is Id1 dependent. *Nature Biotechnol.* **28**, 161–166 (2010).
124. Winkler, I. G. *et al.* Vascular niche E-selectin regulates hematopoietic stem cell dormancy, self renewal and chemoresistance. *Nature Med.* **18**, 1651–1657 (2012).
125. Nikolova, G., Striic, B. & Lammert, E. The vascular niche and its basement membrane. *Trends Cell Biol.* **17**, 19–25 (2007).

Acknowledgements We acknowledge R. Nachman, D. Hajjar and the late A. Marcus and J. Folkman, whose pioneering work in vascular biology inspired us to uncover the instructive functions of endothelial cells. We thank our colleagues W. Schachterle, M. Poulos, K. Shido, S. Rabbany, P. Kermani, R. Lis, M. Seandel, A. Rafii, D. James, J. Scandura, M. Ginsberg, D. Nolan, G. Davis, Z. Rosenwaks, M. Sadelain, I. Riviere and J. Port for critical review and scientific input. We are grateful to Z. Cao for providing expertise in artwork. We apologize to those scientists whose work we could not highlight owing to space limitations. S.R. was supported by the Ansanry Stem Cell Institute, the Empire State Stem Cell Board and New York State Department of Health (C026878, C028117 and C029156), the Howard Hughes Medical Institute, the US National Heart, Lung, and Blood Institute (R01HL115128, R01HL119872 and R01HL128158), the US National Cancer Institute (U54CA163167), the US National Institute of Diabetes and Digestive and Kidney Diseases (R01DK095039) and the Qatar National Priorities Research Program (NPRP 6-131-3-268). J.M.B. was supported by the Ansanry Stem Cell Institute, an American Society of Hematology Scholar Award, an American Federation for Aging Research grant and a Leukemia and Lymphoma Society Quest for Cures award. B.-S.D. was supported by the Ansanry Stem Cell Institute and the American Heart Association (12SDG1213004).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: see go.nature.com/717e9x. Readers are welcome to comment on the online version of this paper at go.nature.com/717e9x. Correspondence should be addressed to S.R. (srafi@med.cornell.edu), J.M.B. (jmb2009@med.cornell.edu) or B.-S.D. (bid2004@med.cornell.edu).

Protein misfolding in the endoplasmic reticulum as a conduit to human disease

Miao Wang¹ & Randal J. Kaufman¹

In eukaryotic cells, the endoplasmic reticulum is essential for the folding and trafficking of proteins that enter the secretory pathway. Environmental insults or increased protein synthesis often lead to protein misfolding in the organelle, the accumulation of misfolded or unfolded proteins — known as endoplasmic reticulum stress — and the activation of the adaptive unfolded protein response to restore homeostasis. If protein misfolding is not resolved, cells die. Endoplasmic reticulum stress and activation of the unfolded protein response help to determine cell fate and function. Furthermore, endoplasmic reticulum stress contributes to the aetiology of many human diseases.

The endoplasmic reticulum (ER) is a specialized organelle that orchestrates the synthesis, folding and transport of at least one-third of the proteins in eukaryotic cells. It regulates protein homeostasis, termed proteostasis, a process that monitors the biogenesis, folding, assembly, trafficking and degradation of all proteins destined for organelles and the extracellular space. These proteins include those that reside in the ER, the Golgi apparatus, lysosomes and the plasma membrane and regulate crucial cellular processes, as well as those that are secreted from cells to mediate intercellular communication. High-quality protein folding is essential for cell survival and function and for normal organismal physiology. In the ER, various control mechanisms exist to ensure anterograde transport of proteins that are correctly folded, modified and assembled. Altered ER homeostasis leads to the accumulation of unfolded or misfolded proteins in the ER lumen, known as ER stress, which activates the unfolded protein response (UPR). UPR activation can promote cell survival, or cell death if ER stress is chronic or severe. We highlight recent advances in understanding of the cellular and molecular mechanisms that trigger ER stress and UPR activation, and their impact on cell fate and function. We also discuss the importance of ER homeostasis for organismal development, physiology and human disease.

Proteostasis in the ER

Protein folding is essential for cell function and survival. In the ER, a number of processes help cells to preserve protein-folding homeostasis.

Protein synthesis, folding and degradation

The dynamic life cycle of a protein begins in the ribosome. Polypeptides generated there that are destined for the secretory pathway usually carry an N-terminal signal sequence that is recognized by the signal recognition particle (SRP), which facilitates their entry to the ER through the translocon complex. Inside the ER, the signal sequence is removed by a signal peptidase and the nascent polypeptides are modified (by N-linked glycosylation, for example), assembled and folded to acquire their native conformation (Fig. 1). Many general and substrate-specific protein disulfide isomerases (PDIs) and peptidyl prolyl isomerases found in the ER increase the rate and efficiency of disulfide bond formation and *cis-trans* isomerization at proline residues. Protein transport through the early secretory pathway is dictated primarily by the chaperone composition of the ER, protein folding status and the structure of the N-linked glycans on the polypeptide. The protein-folding

environment of the ER is distinct from that in the cytosol because of its oxidizing nature, which facilitates the formation of disulfide bonds. It also contains a millimolar concentration of Ca^{2+} that functions as a folding buffer and is essential for chaperone function. Notably, protein folding is the most error-prone step in gene expression.

ER stress

Messenger RNA translation is the stage of gene expression that responds immediately and reversibly to changes in intracellular homeostasis and external stimuli through the covalent modification of initiation factors. Because protein synthesis and secretion rates vary dramatically among cell types and conditions, different levels of nutrients and energy are required to support protein-folding demands. In particular, specialized secretory cells such as the acinar and islet β -cells of the pancreas and plasma cells, as well as malignant cells with high rates of proliferation and protein synthesis, require an ER environment that facilitates efficient folding with quality control. This is supported by a surveillance mechanism that allows only properly folded proteins to move through the secretory pathway. Protein folding is an energy-demanding process that requires coordinated action across organelles and proteins in different cellular compartments, as well as resources from multiple metabolic pathways. Consequently, the efficiency of protein folding in the ER is highly responsive to changes in intracellular environments or extracellular stimuli, especially in specialized secretory cells and malignant cells. Alterations in ER homeostasis can cause the accumulation of misfolded proteins. A wide range of cellular environments and events induce ER stress. These include: increased levels of protein synthesis; impaired ubiquitination and proteasomal degradation; deficient autophagy; energy deprivation; an excess or limitation of nutrients; dysregulated calcium levels or redox homeostasis; inflammatory challenges; and hypoxia. The cause of ER stress varies in different physiological and pathological conditions and will be explored in this Review.

Signal-transduction pathways

The UPR helps the cell to adapt to ER stress. Three UPR signalling pathways have been characterized in metazoans: PERK-like ER kinase (PERK)–eukaryotic translation initiation factor 2 α (eIF2 α), inositol-requiring protein 1 α (IRE1 α)–X-box-binding protein 1 (XBP1), and activating transcription factor (ATF)6 α (Fig. 2). Both PERK and IRE1 α are type I transmembrane proteins with similar ER luminal domain structures and a cytosolic Ser/Thr kinase domain, whereas ATF6 α is

¹Degenerative Diseases Program, Sanford Burnham Prebys Medical Discovery Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, USA.

a type II transmembrane protein that contains a cytosolic cyclic AMP response element-binding protein (CREB)–ATF basic leucine zipper domain.

On acute ER stress, the UPR is activated to restore ER protein-folding homeostasis. This involves the transient attenuation of protein synthesis, an increase in protein folding and transport in the ER and an increase in ER-associated protein degradation and autophagy (Fig. 2). To reduce the protein-folding load on the ER, PERK is activated to phosphorylate eIF2 α at Ser51 (eIF2 α -P), which briefly halts the initiation of mRNA translation. Although it reduces global protein synthesis, eIF2 α -P paradoxically upregulates the translation of many mRNAs, such as *ATF4* mRNA, to increase the capacity for protein transport in the ER, as well as other processes in selected cell types¹. After translation, ATF4 enters the nucleus, where it activates UPR genes that encode proteins that are necessary for the antioxidant response and amino acid biosynthesis and transport. ATF4 also activates the transcription of CCAAT/enhancer-binding protein (C/EBP) homologous protein (CHOP), which forms heterodimers with ATF4 to upregulate genes with functions in the UPR, autophagy and mRNA translation². Once ER protein folding homeostasis is restored, ATF4 and CHOP induce transcription of growth arrest and DNA-damage-inducible protein 34 (GADD34) to direct eIF2 α -P dephosphorylation and restart global mRNA translation, which is essential if cells are to survive an acute insult³.

Both ATF6 α and IRE1 α activate transcriptional pathways that increase the cell's capacity for protein folding, transport and degradation. Under conditions of stress, the activation of IRE1 α , which is encoded by the gene *ERN1*, elicits an endoribonuclease activity that selectively cleaves a 26-nucleotide segment from *XBP1* mRNA. This introduces a translational frameshift that creates transcriptionally active XBP1 (XBP1s), which then enters the nucleus. Genes that are regulated by IRE1 α –XBP1 not only enhance protein folding and transport but also promote degradation pathways — and can thereby resolve protein misfolding. Activation of IRE1 α also mediates mRNA decay to reduce the protein-folding load on the ER, and is termed regulated IRE1-dependent decay (RIDD)⁴. When misfolded proteins accumulate in the ER, ATF6 α transits to the Golgi apparatus, where it is processed to produce a cytosolic fragment that is the primary mediator of the adaptive response to ER protein misfolding. The fragment works mainly by inducing genes that encode protein chaperones to increase the ER's capacity for folding, such as binding immunoglobulin protein (BiP), ER protein 57 (Erp57) and glucose-regulated protein (GRP)94, as well as proteins with ER-associated protein degradation functions.

Emerging evidence suggests that the three UPR signalling branches are not activated simultaneously by ER stress. ATF6 α and IRE1 α activation occurs immediately and is attenuated with time through undefined mechanisms. The activation of PERK follows that of ATF6 α and IRE1 α , however, and persists during chronic ER stress^{5,6}.

ER proteostasis and cell fate and function

Apart from the well-established role of ER stress and the UPR in cell survival and protein secretion, emerging evidence indicates that they have broader functions in the regulation of cell physiology and metabolism.

Cell survival

Once activated, the UPR facilitates cell survival in an attempt to relieve ER stress and restore homeostasis. However, prolonged UPR activation induces apoptosis, mainly through activation of the PERK–eIF2 α –ATF4–CHOP pathway^{5,6}. Studies have demonstrated that CHOP, encoded by the gene *DDIT3*, induces the expression of proapoptotic genes, such as *DR5*, *TRB3*, *BIM* and *PUMA*, and represses the expression of *BCL2*, which triggers apoptosis during ER stress. Moreover, ATF4–CHOP heterodimers initiate the restoration of mRNA translation, which leads to increased protein synthesis, ATP depletion, oxidative stress and cell death². If *Ddit3* is deleted from cells, ER stress causes less protein aggregation in the ER and reduces oxidative stress and apoptosis^{7–10}. This does not occur if any other UPR gene is deleted

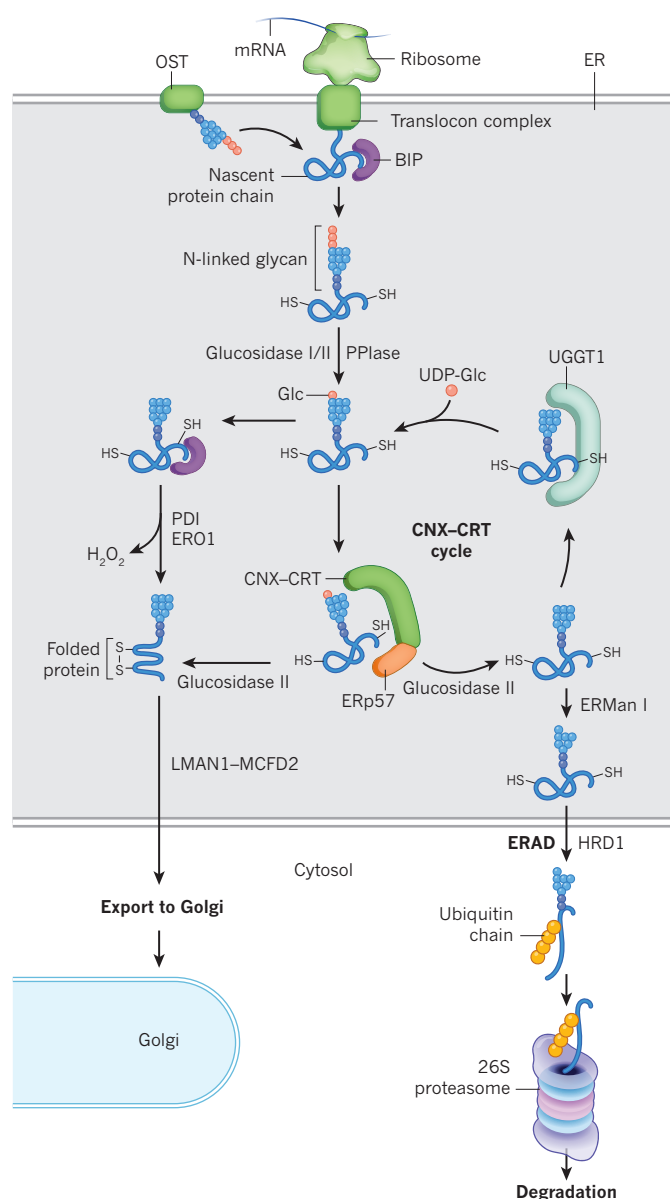


Figure 1 | The dynamic life cycle of a protein in the secretory pathway.

Mature mRNA is loaded onto ribosomes to initiate protein synthesis. Nascent polypeptides enter the ER through the translocon complex in an unfolded state. Oligosaccharyltransferase (OST), a translocon component, is responsible for the transport of such polypeptides. Protein folding in the ER is facilitated by the many resident chaperones, such as PDI, Erp57 and ERO1, as well as peptidylprolyl isomerases (PPIases), which increase the rate of protein folding. Peptide-binding proteins, such as BiP, GRP94 and GRP170, prevent aggregation and actually reduce the rate of protein folding. The addition and trimming of N-linked core oligosaccharides, together with re-glucosylation by UDP-glucose glycoprotein glucosyltransferase 1 (UGGT1), facilitate the interaction of proteins with the chaperones calnexin (CNX) and calreticulin (CRT), which assess the quality of folding and direct transport and degradation of proteins. Correctly folded proteins move to the Golgi apparatus in cargo-trafficking complexes such as LMAN1–MCFD2. At the Golgi they are sorted to different cellular compartments, recycled back to the ER or secreted into the extracellular space (not shown). Unfolded, misfolded or aggregated proteins are retrotranslocated to the cytosol for ubiquitination and degradation by the 26S proteasome through ER-associated protein degradation (ERAD), for which the structure of the N-linked glycan seems to be important. Although proteasome-mediated protein degradation is an important form of disposal for misfolded soluble proteins, the pore-size limitations of the retrotranslocation channel and of the 26S proteasome prevent the degradation of large protein aggregates and complexes that are then subjected to autophagic degradation. ERMan I, ER mannosidase I; HRD1, ERAD-associated E3 ubiquitin-protein ligase HRD1; UDP-Glc, UDP-glucose.

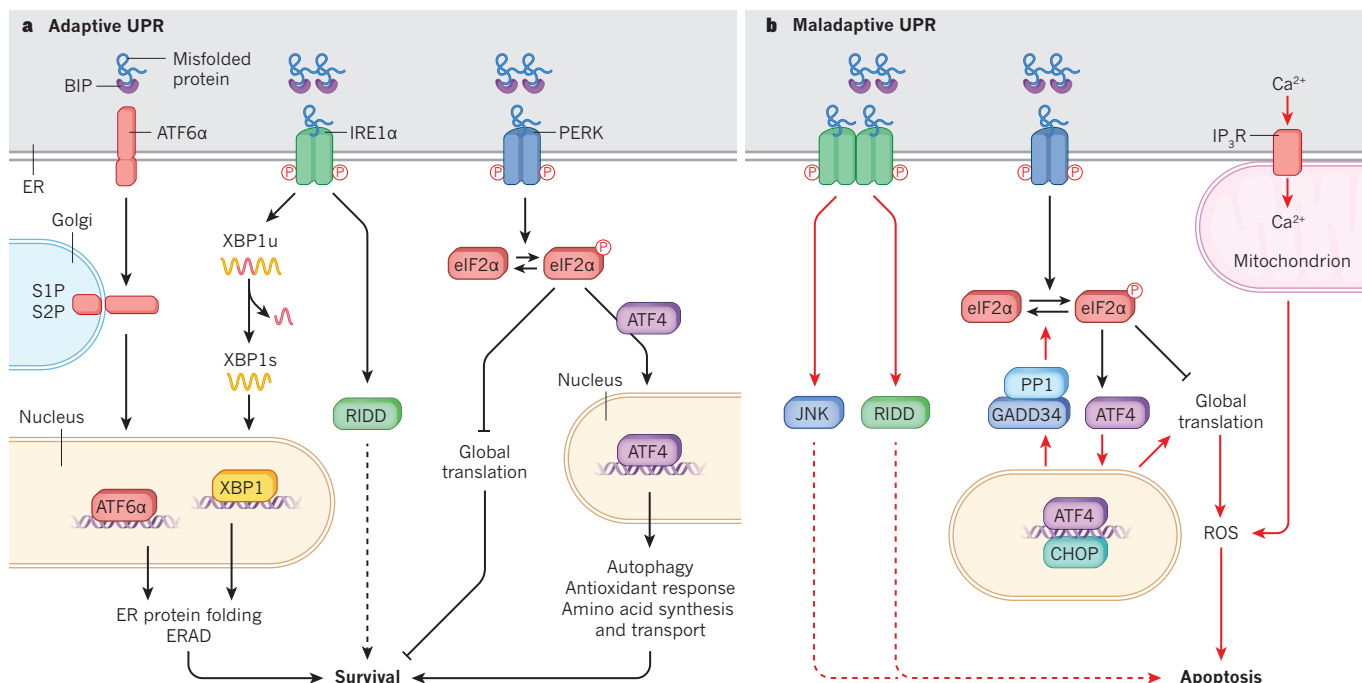


Figure 2 | The UPR signalling pathways. **a**, The adaptive UPR comprises three parallel signalling branches: ATF6α, IRE1α–XBP1 and PERK–eIF2α. The classic view of UPR activation suggests that BiP binds constitutively to the ER-luminal domains of ATF6α, IRE1α and PERK and sequesters them in an inactive form. Misfolded proteins that accumulate in the ER bind to BiP, which causes it to be released from the UPR sensors and triggers their signalling pathways. On release from BiP, ATF6α is trafficked to the Golgi apparatus for processing by the enzymes S1P and S2P, to release a soluble cytosolic fragment that enters the nucleus to induce the expression of target genes. IRE1α and PERK homodimerize or oligomerize and *trans*-autophosphorylate to activate their downstream pathways and promote cell survival. However, this mechanism is challenged by findings that show that the dissociation of

IRE1α or PERK from BiP is not sufficient for UPR activation. In fact, direct interaction between unfolded or misfolded proteins and the UPR sensors might be required^{102,103}. Although the crystal structure of the ER-luminal domain of yeast Ire1p revealed a peptide-binding site that is reminiscent of the major histocompatibility complex (MHC), this groove is not accessible to solvent in the structure of the human IRE1α luminal domain¹⁰⁴, which suggests that the process is more complex. The mechanisms that underlie differential regulation of the dynamics and duration of UPR activation remain unclear. **b**, The maladaptive UPR is induced by sustained activation of the PERK pathway, which is the result of prolonged severe ER stress and leads to apoptosis (solid red lines). The role of IRE1α-induced JNK and RIDD in ER-stress-induced apoptosis remains unclear (Box 1) (dashed lines).

instead. CHOP also activates ER oxidase 1α (ERO1α), an oxidoreductase that mediates the transfer of electrons during disulfide bond formation to molecular oxygen to produce hydrogen peroxide. This reaction augments the production of reactive oxygen species (ROS) and inositol-1,4,5-trisphosphate receptor (IP₃R)-mediated Ca²⁺ efflux from the ER. Ca²⁺ that is released from the ER is taken up by mitochondria through the mitochondria-associated ER membrane, which promotes mitochondrial production of ROS through different mechanisms. These include the release of cytochrome *c*, stimulation of Krebs cycle dehydrogenases and activation of nitric oxide synthase. Flux of Ca²⁺ between the ER and mitochondria probably couples the ATP demands of protein folding with mitochondrial bioenergetics for ATP production. Thus, ER stress induces oxidative stress and impairs mitochondrial function¹¹, which leads to cell death in a CHOP-dependent manner.

Cell growth and differentiation

Cell proliferation and differentiation entail an increase in protein synthesis that primes cells for ER stress and UPR activation. During ER stress, surveillance machinery delays inheritance of the ER and the process of cytokinesis that is dependent on the mitogen-activated protein kinase (MAPK) SLT2 but not on PERK, IRE1α and ATF6α — the sensors of UPR. In budding yeast, in the absence of Slf2, ER stress is transmitted to daughter cells, which results in the death of both mother and daughter cells¹².

The role of the UPR in cell differentiation was first demonstrated by the requirement for the IRE1α–XBP1 pathway in the differentiation of plasma cells¹³. Although early studies suggested that IRE1α, but not the splicing of XBP1, was required for immunoglobulin gene rearrangement, these findings have not been confirmed in conditional

gene knockout mice¹⁴. More importantly, the deletion of either *Ern1* or *Xbp1* in mice produces the same defect in the differentiation of mature B cells into plasma cells^{13,14}. Because cell differentiation is associated with a sixfold expansion of the ER, it is tempting to speculate that the IRE1α–XBP1 pathway is required for expansion of the secretory pathway in cells that secrete large amounts of protein^{13–15}. Interestingly, activation of the IRE1α–XBP1 pathway does not result from an increased secretory protein load from immunoglobulins. Instead, it is caused by a differentiation-dependent signal from B-cell receptors that upregulates genes encoding components of the secretory pathway¹⁵ and plasma-cell transcription factors^{16,17}, such as Mist1 (ref. 18). Surprisingly, deletion of PERK, eIF2α-P or ATF6α did not generate defects in plasma-cell differentiation, which provides convincing evidence that these UPR pathways have different functions in different cell types. XBP1 can induce a wide spectrum of secretory pathway genes, as well as an increase in ER and lysosome protein content, mitochondrial mass and function, ribosome number and levels of protein synthesis. Therefore, the IRE1α–XBP1 pathway contributes markedly to the characteristic phenotype of professional secretory cells^{15,18} such as gastric zymogenic cells¹⁹, β-cells^{17,20} and intestinal Paneth cells²¹. Moreover, the IRE1α–XBP1 pathway is involved in the development of other immune cells and functions, including CD8⁺ T-cell differentiation during acute infection²² and dendritic-cell development and function²³. Interestingly, studies suggest an XBP1-independent role for IRE1α in cell differentiation. For example, Ire1α-mutant photoreceptors, but not Xbp1 mutants, showed defects in differentiation and rhabdome morphogenesis in *Drosophila melanogaster*, possibly through the regulation of RIDD²⁴. IRE1α also induces the transdifferentiation (lineage reprogramming) of germ cells into ectopic somatic cells in *Caenorhabditis elegans*, whereas the XBP1

and other UPR pathways did not show similar effects²⁵. Further studies are needed to reveal the underlying mechanisms.

The PERK pathway also contributes to cell differentiation. For example, when phosphorylated by p38 MAPK, CHOP inhibits adipocyte differentiation *in vitro*³⁶. Furthermore, both *in vitro* and *in vivo* adipocyte differentiation are blocked by CHOP expression after PERK activation and eIF2 α phosphorylation²⁷. This is consistent with the idea that CHOP inhibits the C/EBP family of transcription factors, which are required for adipocyte differentiation²⁷. In contrast, PERK harbours intrinsic lipid kinase activity that phosphorylates diacylglycerol to generate phosphatidic acid²⁸, which could offer an explanation for how PERK promotes adipocyte differentiation through the induction of lipogenic pathways²⁹. Moreover, ATF4 promotes adipocyte differentiation *in vitro* by directly activating C/EBP- β and peroxisome proliferator-activated receptor γ (PPAR γ)³⁰. This discrepancy might result from the temporal pattern of PERK–eIF2 α -P activation that leads to the induction of different sets of genes at different times after ER stress: the PERK–eIF2 α pathway provides cell-survival signals at early times of ER stress and apoptosis signals at later times⁵. Further kinetic studies of physiological inducers of ER stress are needed to dissect the requirement for the PERK–eIF2 α pathway and its downstream targets at different stages of adipocyte differentiation. The PERK–eIF2 α pathway is also involved in the differentiation of other cell types, such as epithelial-to-mesenchymal transition³¹ and vascular smooth-muscle cell proliferation³².

Cellular metabolism

The link between ER stress signalling and metabolism was highlighted when researchers found that humans and mice with PERK deletions and mice with mutations at the PERK phosphorylation site in eIF2 α have a defect in glucose metabolism^{33–35}. ER homeostasis and UPR activation are now known to be central to both glucose and lipid metabolism. Glucose homeostasis is controlled tightly by insulin and glucagon levels in the blood. Insulin-secreting β -cells and hepatocytes that are responsive to insulin and glucagon have important roles in maintaining glucose homeostasis. The deletion of UPR components, such as Perk³⁶, eIF2 α -P^{35,37}, IRE1 α ¹⁷, Xbp1 (ref. 20), Atf6 α ³⁸ and p58^{IPK} (ref. 39), in β -cells of adult mice reduced insulin synthesis and secretion, increased β -cell death and caused hyperglycaemia, with or without a metabolic challenge. Notably, mice with a mutation in eIF2 α -P, an *Ern1* deletion or a p58^{IPK} deletion^{17,37,39} that were fed a diet enriched with antioxidants showed improved β -cell function. In contrast, deletion of either *Ddit3* (ref. 9) or *Atf4* (ref. 40) prevented the generation of ROS and protected mice from hyperglycaemia that would have been induced by a high-fat diet and β -cell dysfunction. *Ddit3* deletion dramatically protected β -cells in models of type 2 diabetes, including *db/db* mice, by reducing oxidative damage⁹. However, *Ddit3*^{−/−} mice did not exhibit any marked phenotype in the absence of a stress signal. These observations are consistent with the finding that ATF4 requires CHOP to cooperatively induce oxidative stress and cell death². Glucagon-mediated gluconeogenesis in hepatocytes also plays an important part in glucose homeostasis, although the roles of ER stress and the UPR are less clear. Inhibition of either eIF2 α -P or IRE1 α signalling in the liver decreases hepatic glucose production in mice. This could be attributed to the reduced expression of genes that encode gluconeogenic enzymes, which reduces basal plasma glucose^{41,42}. ATF4 also augments gluconeogenesis by controlling the transcriptional activity of forkhead box protein O1 (FOXO1)⁴³. In contrast, ATF6 α seems to reduce hepatic gluconeogenesis by inhibiting activation of CREB and expression of gluconeogenic genes⁴⁴. However, mice with a whole-body deletion of *Atf6a* did not exhibit a marked metabolic phenotype⁴⁵, and it has been demonstrated that the only detectable phenotype in humans with hypomorphic mutations in *ATF6a* is age-related degeneration of the retinal cone photoreceptors⁴⁶. The overall requirement for ATF6 α in metabolism therefore remains a mystery. A more important question is why loss of function in this general transducer of the UPR pathway causes such a selective, cell-type-specific defect, in which retinal rod cells survive

but the cones do not. ER stress and UPR pathways also contribute to glucose homeostasis in other tissues.

The exposure of cells to elevated levels of free fatty acids, such as palmitate and stearate, induces ER stress⁴⁷ through multiple mechanisms. These include: deregulated protein palmitoylation; aberrant accumulation of ROS; alterations in membrane fluidity⁴⁸; elevated levels of fatty-acid oxidation; and an increased protein-folding load that results from hyperactivation of the mammalian target of rapamycin (mTOR) anabolic signalling pathway. In response to physiologically or pathologically induced ER stress, the UPR pathways PERK–eIF2 α and IRE1 α –XBP1 can activate C/EBPs and sterol regulatory element-binding protein (SREBP)1 and 2, which are important transcriptional activators of fatty-acid and cholesterol synthesis⁴⁹, to accommodate the need for ER expansion. SREBP-mediated lipogenesis also maintains the balance between the saturated and monounsaturated fatty-acid pools to prevent lipotoxicity in tumour cells⁵⁰. However, the interpretation of these findings is complicated by the observation that severe and persistent ER stress dampens lipogenesis and increases fatty-acid oxidation while causing the accumulation of lipids in the liver⁵¹. Each UPR subpathway seems to maintain lipid homeostasis through a different mechanism.

Protein secretion

The ER is responsible for the post-translational modification, folding and transport of secretory proteins such as cytokines and hormones. ER stress blunts the synthesis and secretion of secretory proteins through multiple mechanisms. A well-studied example is insulin. In β -cells, proinsulin interacts with many ER proteins to facilitate its folding and transport⁵². Ablation of PERK activity in human and rodent β -cells impairs the synthesis of insulin^{36,37} and attenuates its secretion owing to alterations in store-operated Ca²⁺ entry and sarcoplasmic–ER Ca²⁺ ATPase activity⁵³. Embryonic ablation of XBP1 markedly impairs the processing of proinsulin owing to increased ER stress and IRE1 α hyperactivation-induced RIDD, which degrades mRNAs that encode proinsulin processing enzymes²⁰. In contrast, *Ern1* deletion in mature β -cells impairs proinsulin mRNA translation — without altering levels of insulin mRNA — by downregulating expression of genes required for the SRP, the SRP receptor, the translocon complex and the signal peptidase complex¹⁷. Thus, insulin levels can be affected by ER stress and UPR activation at the transcription, translation and secretion stages. Similarly, ER stress and UPR pathways influence other secretory proteins at the transcriptional or translational level, such as lysozyme in Paneth cells^{21,54} and cytokines in macrophages⁵⁵. Because these secretory proteins affect the function of distant organs, ER stress not only impairs the survival and function of secretory cells but also has an effect on the organism as a whole.

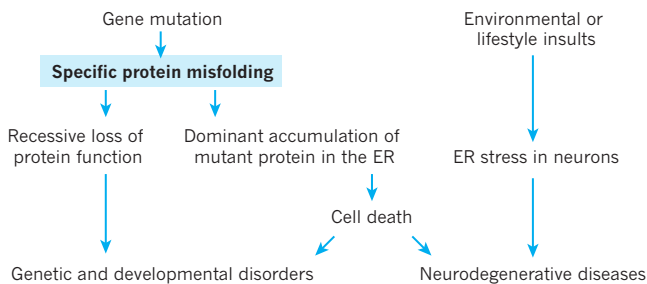
ER stress in human disease

ER stress *in vivo* is induced by both intrinsic and extrinsic factors. The most common physiological intrinsic factor that leads to ER stress is an increase in protein synthesis, which occurs during growth factor stimulation, cell proliferation, differentiation or senescence. Gene mutations that cause proteins to misfold and defects in UPR pathways can also induce ER stress directly. More often, protein misfolding is caused by changes in the environment, such as inflammation, hypoxia or nutrient deprivation. Because all of these factors are involved in pathogenesis, it is unsurprising that ER stress contributes to the initiation and progression of numerous diseases in humans (Fig. 3).

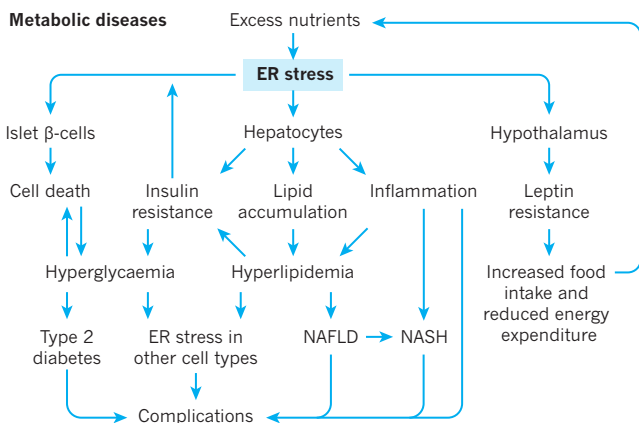
Genetic diseases

Deletions in *Ern1* or *Xbp1* are lethal at the embryonic day (E)12 stage of mouse fetal development, whereas deletions in *Perk* or eIF2 α -P through mutation of the phosphorylation site in *Eif2a* are lethal after birth. Deletion of *Atf6a* has no apparent phenotype⁴⁵ unless challenged with ER stress, when mice cannot mount an adaptive response. And mice in which both *Atf6a* and its homologous gene *Atf6b* are deleted

Genetic disorders and neurodegenerative diseases



Metabolic diseases



Inflammatory diseases

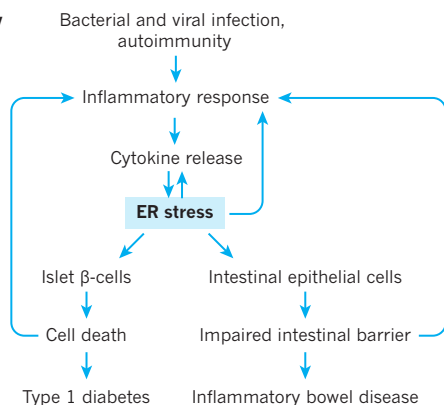


Figure 3 | ER stress and human disease. ER stress and UPR activation have important roles in the pathogenesis of many diseases. The induction of ER stress *in vivo* is attributable to both intrinsic and extrinsic factors. For example, gene mutations can cause defects in the folding and transport of certain proteins. The loss of function and toxicity that is associated with the accumulation of mutant protein in the ER leads to genetic and developmental disorders and neurodegenerative diseases. Environmental or lifestyle insults that trigger ER stress in neurons also contribute to the development of neurodegenerative diseases. More often, however, environmental insults such as an excess of nutrients or inflammation induce protein misfolding in the ER, which compromises cell function and triggers cell death and the development of metabolic and inflammatory diseases. NASH, non-alcoholic steatohepatitis.

exhibit early embryonic lethality⁵⁶. Thus, although ATF6β seems not to regulate UPR genes, ATF6α and ATF6β provide complementary yet unknown functions in early development. These findings indicate that each UPR pathway has unique roles in development⁵⁷. Exome sequencing has identified many mutations in ER components in humans that result in cell-type-specific deficiencies. For example, mutations in *PERK* cause β-cell failure and early-onset diabetes owing to reduced levels of eIF2α-P^{33,58}. A mutation in *Si1*, an adenine nucleotide exchange factor for BiP, disrupts ER function and causes neurodegeneration⁵⁹. Mutations in many secretory proteins, such as insulin, α-1 antitrypsin and

clotting factors, cause misfolding that results in either recessive genetic disorders due to loss of protein function⁵⁹ or dominant genetic disorders due to toxicity⁶⁰ that are associated with accumulation of the mutant protein in the ER (Fig. 3). In the rare bleeding disorder combined deficiency of factor V and factor VIII (F5F8D), the secretion of the blood-clotting factors V and VIII is reduced to about 30% of the normal level, but there is no defect in the production of other blood-plasma proteins. Two-thirds of patients with F5F8D have loss-of-function mutations in the gene *LMAN1* and *LMAN1* (also known as ERGIC-53), the mannose-binding lectin that it encodes; the other one-third have mutations in the gene *MCFD2*, which encodes the Ca²⁺-binding multiple coagulation factor deficiency protein 2 that interacts with *LMAN1*. The *LMAN1*–*MCFD2* heterodimer interacts with cargo to transport protein from the ER to the *cis*-Golgi apparatus⁶¹. The clotting factors V and VIII contain heavily glycosylated central regions, which leads to unique requirements for protein folding and transport. Other gene mutations that lead to protein misfolding and genetic or developmental disorders in humans are summarized in Supplementary Table 1.

Neurodegenerative disease

Age-related diseases are commonly associated with the accumulation of misfolded and aggregated proteins and with oxidative stress. Most evident is the accumulation of misfolded proteins in neurons in progressive neurodegenerative diseases, which contributes directly to ER-stress-induced neuronal apoptosis⁶². Specific misfolded proteins that are induced by gene mutations have been identified in neurodegenerative diseases, such as Parkinson's disease and Alzheimer's disease (Supplementary Table 1). A prominent pathological hallmark of Alzheimer's disease is the abnormal accumulation of the amyloid-β (Aβ) peptides in the brain, which induces neuronal loss and cognitive deficits owing to both ER stress and oxidative stress⁶³. Elevated levels of eIF2α-P and ATF4 have also been observed in the brains of people with Alzheimer's disease and in mouse models of the disease⁶⁴. An increase in ER stress in brains that are affected by this disease can also result from the S-nitrosylation of PDI or IRE1α, which compromises their function^{65,66}. Accumulation and aggregation of the protein α-synuclein (α-syn) defines the pathobiology of Parkinson's disease. Indeed, ER stress was identified as an early pathogenic phenotype in neurons derived from induced pluripotent stem cells from patients with α-syn mutations⁶⁷. α-Syn inhibits ATF6α processing directly through physical interactions, and it also indirectly impairs ER–Golgi protein transport, which enhances ER stress and apoptosis⁶⁸. Mutations in the E3 ubiquitin ligase parkin also cause the accumulation and aggregation of α-syn owing to reduced ubiquitin ligase activity⁶⁹. Many studies in mouse models have demonstrated the important role of UPR pathways in preventing neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, and in protecting against spinal-cord injury caused by genetic mutation or physical or chemical insults⁶². For example, deletion of *Xbp1* in the brain was shown to improve symptoms of Huntington's disease — presumably through the activation of autophagy to degrade aggregated proteins⁷⁰. However, mutations in UPR pathways that lead to neurodegeneration have not been identified in humans.

Metabolic disease

Genetic disruption of UPR pathways in mice uncovered a role for ER homeostasis in the regulation of glucose and lipid metabolism. ER stress and UPR pathways contribute to the pathogenesis of metabolic diseases such as diabetes⁷¹ and non-alcoholic fatty liver disease (NAFLD)⁷² (Fig. 3), and both type 2 diabetes and NAFLD are associated with an excess of nutrients in the diet. Hyperlipidaemia causes ER stress in hepatocytes, and studies^{72–75} indicate that the UPR is essential to limit the accumulation of lipids in these cells. In mice, ER stress induced by injecting tunicamycin causes lipid accumulation in hepatocytes, which peaks after 24 hours and disappears over several days^{51,76}. Interestingly, the clearance of lipids requires all three sensors of UPR⁵¹. Therefore, ER

stress promotes NAFLD probably by enhancing lipogenesis, increasing lipolysis, reducing the secretion of very-low-density lipoproteins (VLDLs) and decreasing fatty-acid oxidation^{72,75}. Hyperlipidaemia also induces ER stress in the hypothalamus of the brain, which leads to leptin resistance and UPR-pathway activation in hepatocytes⁷⁷. ER stress in hepatocytes causes insulin resistance to develop through hyperactivation of the c-Jun N-terminal kinase (JNK) pathway and subsequent serine phosphorylation of insulin receptor substrate 1 (IRS1)⁷¹. Insulin resistance places a heavy burden on β -cells to synthesize and secrete more insulin. If UPR pathways cannot adapt to the increase in proinsulin synthesis, β -cell apoptosis occurs. A loss of 90% of pancreatic islet β -cell mass leads to the hyperglycaemia that is associated with type 2 diabetes. In late-stage type 2 diabetes in humans, β -cells accumulate aggregates of islet amyloid polypeptide (IAPP)⁷⁸ — perhaps because of the hostile protein-folding environment induced by long-term ER stress in β -cells. IAPP might also augment ER stress, which would increase the occurrence of both β -cell apoptosis and severe

hyperglycaemia. However, the importance of IAPP aggregates in β -cell failure is unknown. In type 1 diabetes, an autoimmune disease, the inflammatory environment induces ER stress in β -cells, which might impair insulin biosynthesis and secretion and lead to β -cell apoptosis⁷⁹. Chronic exposure to the elevated glucose, lipid or cytokine levels that are associated with diabetes also leads to ER-stress-induced β -cell toxicity. Indeed, islet cells from individuals with type 1 or type 2 diabetes contain increased levels of CHOP^{78,80}, which indicates the presence of ER stress and apoptosis. Hyperlipidaemia and hyperglycaemia also induce ER stress in other cell types, such as endothelial cells and neurons, and contribute to the development of complications in diabetes and NAFLD. Notably, hypomorphic mutations in the gene *CREB3L3* — which encodes CREB-H and is related to *ATF6A* — causes severe hypertriglyceridaemia in humans⁸¹. CREB-H was originally identified as a hepatocyte-specific transcription factor that is induced at the transcriptional level by inflammatory stimuli and then, similarly to *ATF6A*, regulated by site-1 protease (S1P)- and site-2 protease (S2P)-mediated

BOX 1

Outstanding questions about ER stress

● Does regulated RIDD contribute to ER-stress-induced apoptosis?

Sustained IRE1 α activation might lead to cell death through RIDD, which regulates gene expression by modulating the production and stability of microRNA. It also mediates the degradation of selective mRNAs that are associated with the ER membrane. Therefore, RIDD can deplete essential components of the protein-folding machinery to exacerbate ER stress and cell death. It remains unclear whether this process depends solely on the dimerization¹¹¹ or oligomerization status of IRE1 α ¹¹⁰. RIDD targets seem to differ according to cell type, although biogenesis of lysosome-related organelles complex-1 subunit 1 (BLOC1S1) is probably a universal RIDD substrate¹¹². RIDD might also augment apoptosis by reducing the expression of microRNAs such as miR-17, miR-34a, miR-96 and miR-125b, which repress caspase 2 expression¹¹³. However, the importance of caspase 2 activation in ER-stress-induced apoptosis is in doubt¹¹⁴. In contrast, RIDD protects cancer cells against cell death by degrading their *DR5* mRNA¹¹⁵ and hepatocytes by degrading their *Cyp1a2* and *Cyp2e1* mRNAs¹¹⁶. Other reports show that RIDD has no significant impact on cell survival or apoptosis^{111,112}. These results call into question the physiological importance of RIDD in cell survival. RIDD might have cytotoxic or cytoprotective effects depending on cell type or cellular context.

● Does IRE1 α regulate JNK activation to cause ER-stress-induced apoptosis?

A further link between IRE1 α and cell death has been proposed: IRE1 α interacts with the adaptor protein TRAF2 to activate JNK, and *Ern1*^{-/-} mouse embryonic fibroblasts (MEFs) display reduced levels of JNK activation. However, direct interaction between IRE1 α and JNK and subsequent JNK phosphorylation have not been demonstrated. In addition, deletion of *Xbp1* leads to IRE1 α hyperactivation¹¹⁷ and activation of the MAPK ASK1 and its downstream target JNK, although the mechanism for JNK activation remains unclear and could involve production of ROS that inhibit MAPK phosphatases¹¹⁸. However, IRE1 α hyperactivation has also been shown to reduce JNK activation in hepatocytes on acetaminophen (paracetamol) challenge¹¹⁶. Therefore, there is no convincing evidence yet for IRE1 α -mediated activation of JNK.

● Does PERK have other substrates that influence cell survival, function and differentiation? It is unlikely that PERK has more substrates than just eIF2 α . Early studies suggested that PERK phosphorylates NRF2, which triggers dissociation of the NRF2-Keap1 complex to stabilize NRF2 and activate transcription of the antioxidant response¹¹⁹. However, PERK-dependent changes in gene expression

were prevented by Ser51Ala mutation in eIF2 α ¹²⁰, which suggests that additional pathways of transcription activation following eIF2 α phosphorylation do not exist. Thus, the physiological importance of alternative PERK substrates remains in doubt.

● Are the induction of ATF4 and CHOP or the splicing of XBP1 mRNA indicators of ER stress?

Many studies measure the induction of ATF4 and CHOP to estimate ER stress. However, these might not be accurate indicators because — in addition to PERK — three other eIF2 α kinases, GCN2, HRI and PKR, can increase levels of ATF4 and CHOP through eIF2 α phosphorylation. The phosphorylation of eIF2 α might also be regulated by protein phosphatase 1C through its regulatory subunits GADD34 and CREP. Furthermore, IRE1 α can be activated in an ER-stress-independent manner through interaction of TLR2 or TLR4 (refs 55, 83) with the E3 ligase TRAF6. This prevents the protein phosphatase 2A-mediated dephosphorylation of IRE1 α and IRE1 α ubiquitination and degradation¹²¹. Therefore, neither the induction of ATF4 and CHOP nor the splicing of *XBP1* mRNA (an event downstream of IRE1 α activation) reliably indicate the presence of ER stress.

● How do small molecules regulate the phosphorylation of PERK-eIF2 α ?

The efficacy of modulating the PERK-eIF2 α pathway has been demonstrated in various mouse models of disorders, such as neurodegeneration. Whereas a reduction in eIF2 α phosphorylation rescues synaptic deficits and neuronal loss in mice with prion disease by restoring protein synthesis^{122,123}, the enhancement of eIF2 α phosphorylation by inhibiting GADD34 also provides neuroprotection by blocking protein synthesis¹⁰⁷. It will be necessary to fine-tune inhibition of the PERK-eIF2 α pathway with respect to cell type to obtain therapeutic effects¹²⁴. Furthermore, caution must be taken when interpreting and translating results because the inhibitors might act through off-target effects *in vivo*. Indeed, the frequently used PERK inhibitor GSK 2656157 was shown to induce apoptosis in PERK-deleted MEFs¹²⁵.

● How should ER stress that is triggered by pharmacological

induction of the UPR be interpreted? Although pharmacological agents such as thapsigargin, tunicamycin, dithiothreitol, proteasome inhibitors and brefeldin A provide a convenient approach to activating the UPR, these insults have multiple effects that might not be a consequence of ER protein misfolding. It would be better to measure the cellular response to a specific defect in protein misfolding. In addition, different misfolding defects might activate different downstream responses.

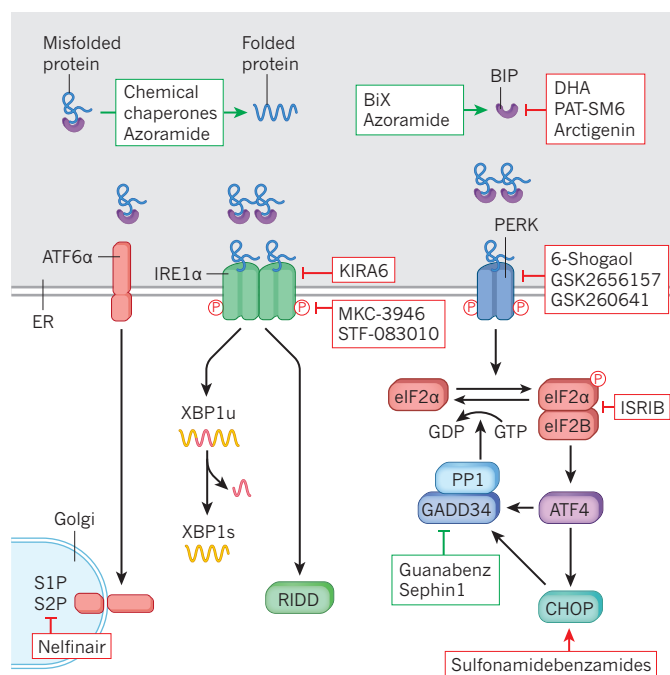


Figure 4 | ER stress and UPR pathways as therapeutic targets. Many small molecules have been identified that alleviate (green boxes) or induce (red boxes) ER stress through different mechanisms. Chemical chaperones, the antidiabetic compound azoramide¹⁰⁵ and BiP inducer X (BiX)¹⁰⁶ alleviate ER stress by enhancing the folding capacity of the ER. Guanabenz and Sephin1 (ref. 107) inhibit GADD34 to promote persistent eIF2α phosphorylation, which reduces mRNA translation and ER stress. Small molecules that block UPR activation⁸² include integrated stress response inhibitor (ISRIB), which targets the interaction between eIF2α and eIF2B to prevent the effects of eIF2α phosphorylation^{108,109}, and KIRA6, which inhibits the RNase activity of IRE1α by dissociating IRE1α oligomers to prevent RIDD¹¹⁰. Other small molecules shown here that enhance ER stress are reviewed in ref. 82.

intramembrane proteolysis on transport to the Golgi apparatus during ER stress⁷³. However, CREB-H does not activate UPR genes. Instead, it induces the expression of genes involved in the systemic inflammatory response (that encoding C-reactive protein) and lipid homeostasis (those for apolipoproteins and lipoprotein lipase)⁸¹.

Inflammation

Signalling crosstalk has been shown to occur between ER stress and inflammatory responses^{73,82,83}. Pro-inflammatory stimuli, such as ROS or Toll-like receptor (TLR) ligands, trigger ER stress and activate the UPR to initiate or amplify inflammatory responses. Elevated inflammatory cytokines in diabetic pancreatic islets trigger ER stress in the β -cells⁸⁴. And IRE1α can be activated by TLR signalling in the absence of detectable ER stress^{55,83} (Box 1). In the setting of obesity, inflammatory input leads to the S-nitrosylation of IRE1α through increased activity of inducible nitric oxide synthase in hepatocytes, which compromises IRE1α activation and causes ER stress⁶⁶. During ER stress in hepatocytes, activated ATF6α and CREB-H form homo- or heterodimers to enhance the expression of acute-phase response genes, including those that encode C-reactive protein and serum amyloid P component, which in turn amplify inflammation⁷³. NF- κ B, a master transcriptional regulator of the pro-inflammatory response, can be activated by all three UPR pathways through the regulation of its inhibitor I κ B^{82,85}. Moreover, the IRE1α–TNF receptor-associated factor (TRAF)2 complex can recruit and activate JNK, which leads to increased expression of pro-inflammatory genes through enhanced activity of the transcription factor AP-1 — although the importance of this pathway remains in question (Box 1). ER stress also can trigger activation of the NLRP3 inflammasome in both β -cells and macrophages^{86,87}, possibly through ER-stress-induced mitochondrial

damage⁸⁸. Therefore, perturbed ER homeostasis in parenchymal cells can trigger and augment inflammatory responses, which leads to inflammatory disease. For example, in inflammatory bowel disease, inflammatory cytokines can induce ER stress in intestinal epithelial cells⁸⁹, which in turn augments the inflammation (Fig. 3). Furthermore, ER stress in Paneth cells reduces lysozyme synthesis and secretion^{21,54}, which impairs the host defence machinery and contributes to the development of inflammatory bowel disease. Such findings indicate that extreme crosstalk exists between signalling pathways and therefore needs further study.

Cancer

ER stress and UPR activation have been documented in a variety of human cancers. Because cancers usually arise and progress in a stressful microenvironment, transformed cells might use UPR activation as a survival strategy. In fact, many studies have demonstrated a crucial role for the UPR in both tumour cells and stromal cells in the tumour microenvironment as cancer develops and progresses⁸². UPR pathways might also interact with oncogenes or tumour-suppressor genes to modulate their function in the development of cancer. Hyperactivation of oncogenes or loss-of-function mutations in tumour-suppressor genes often lead to an increase in protein synthesis, which results in an increased protein-folding load in the ER. Consequently, the UPR is activated to increase the capacity for protein folding⁸². During malignant progression, cancer cells activate pathways that affect cells in the tumour microenvironment, such as immune cells and endothelial cells, to support tumour growth. That might require UPR signalling to increase the folding and secretion of cytokines, angiogenesis factors and extracellular matrix components. Rapidly proliferating cancer cells often encounter a hostile environment that disrupts protein folding in the ER and activates the UPR⁸². Except for intensive studies on tumour growth, the role of ER stress in the initiation of tumours remains unclear. Interestingly, ER stress has been shown to generate pre-oncogenic cells that lead to hepatocellular carcinoma under conditions of a high-fat-diet-induced inflammatory environment⁹⁰, indicating an oncogenic role for ER stress.

Therapeutic targets

ER stress is implicated in the aetiology of many diseases. Therefore, UPR pathways could be important therapeutic targets for modulating ER stress and associated diseases.

Reduction of ER stress

Chemical chaperones are small molecules that prevent protein aggregation, facilitate protein folding and reduce ER stress — both *in vitro* and *in vivo* — by stabilizing protein-folding intermediates (Fig. 4). The chemical chaperone tauroursodeoxycholic acid has been studied intensively as a treatment for many diseases. Its therapeutic potential has been demonstrated in mouse models of diabetes⁹¹ and inflammatory bowel disease⁸⁹. In Europe, the chaperone has been approved as a therapy for cholelithiasis and cholestatic liver disease. Clinical studies have shown that treatment with the chaperone for 1 to 6 months is safe and efficacious in patients with disorders such as primary biliary cirrhosis, liver cirrhosis, hepatitis C virus (HCV)-related chronic hepatitis and insulin resistance^{92,93}. Several drugs approved by the US Food and Drug Administration (FDA) also achieve their therapeutic effects partially through the restoration of ER homeostasis. For example, a dual PPAR agonist used in the treatment of diabetes reduced ER stress in the liver and adipose tissue of *db/db* mice and subsequently increased sensitivity to insulin⁹⁴. And the glucagon-like peptide 1 receptor (GLP-1R) agonist exendin-4 reduced ER stress in pancreatic islet β -cells *in vivo* by enhancing the expression of ATF4 (ref. 95). Selective pharmacological chaperones that are protein specific have been developed to facilitate the folding and transport of various proteins that are misfolded as a result of gene mutations⁹⁶, such as G-protein-coupled receptors, the Δ 508Phe mutant cystic fibrosis transmembrane conductance regulator

(CFTR) and lysosomal enzymes. ER stress might also be alleviated *in vivo* by gene therapy. For example, delivery of an active form of *Xbp1* using an adeno-associated vector significantly reduced the accumulation of mutant proteins in neurons in a mouse model of Huntington's disease⁹⁷.

Induction of ER stress for cancer therapy

Owing to the important role of ER stress and UPR pathways in cancer, the induction of ER stress represents a new therapeutic strategy for killing tumour cells⁸². The selective advantage of inhibiting the UPR in tumour cells is that these cells are often 'addicted' to UPR pathways to survive stressful microenvironments. Such therapies can induce severe ER stress that leads to cell death. Moreover, ER stress in the tumour microenvironment modulates the function of cancer-supporting stromal cells, such as endothelial cells, and suppresses tumour growth. Because UPR-pathway activation has both pro- and anti-survival effects on cells, caution is necessary when designing therapies that target UPR components and when interpreting the results. The proteasome inhibitor bortezomib is the first FDA-approved drug for cancer therapy that functions as an inducer of ER stress. Some candidate compounds are designed to inhibit specific components of the protein-folding or UPR pathways, such as PERK, IRE1 α , HSP90, GRP94 and BiP, thereby augmenting ER stress and cell death in cancer cells (Fig. 4). The significant correlation between BiP expression, tumour stage and drug resistance has led to the testing of a number of anti-BiP therapies in a variety of cancers⁸². Alternatively, therapeutic agents that activate the PERK–ATF4–CHOP pathway can be used to enhance the apoptotic UPR in cells with constitutive UPR activation^{31,98}. Inhibition of the UPR or activation of the PERK pathway could sensitize cancer cells and cancer stromal cells to other therapies, which offers an advantage for combination therapies⁸². Therefore, such therapies that include drugs that target ER stress and UPR pathways represent one of the most promising anticancer approaches.

Future directions

Although we have gained tremendous insight into the mechanisms of protein folding and the adaptive responses to protein misfolding in the ER that regulate cell fate, function and disease in humans, outstanding questions remain (Box 1). Of note, most studies use measurements of UPR activation and Ca²⁺ or ROS levels as indirect markers for ER stress. However, these parameters can be altered by other stress conditions too, and even UPR pathways can be activated independently of ER stress⁸² (Box 1). Therefore, changes in these parameters do not always demonstrate ER stress. Technologies must be developed that monitor the status of protein folding *in vivo*. Many methods that were developed to directly detect misfolded proteins or aggregates *in vitro* have had limited success *in vivo*. Although electron microscopy can detect ER dilation and protein aggregates, the severity of protein misfolding that occurs in human disease — and which proteins misfold under these circumstances — remains unclear. Thus, the mechanisms that activate and sustain UPR signals during the pathogenesis of diseases are not fully understood. Furthermore, because UPR pathways are generally activated only transiently after ER stress, the temporal aspects of UPR activation must also be considered. Detailed kinetic studies should be done on the activation of UPR pathways. In addition, age-related degenerative diseases are generally associated with the accumulation of both protein aggregates and oxidation products. It is important to understand whether ER stress is the cause or the effect. The UPR is probably not cell autonomous, and signals could be transmitted from one cell to another, sometimes across considerable distances. For example, tumour cells with ER stress secrete soluble factors that can initiate ER stress and upregulate the expression of pro-inflammatory cytokines in macrophages⁹⁹. Forced expression of active XBP1 in neurons can activate the UPR in intestinal¹⁰⁰ and liver tissues¹⁰¹. Although these findings are intriguing, the precise transmitters remain unclear. Identification of the transmitters and mechanisms that activate and prolong

UPR signalling in disease will help us to further our understanding of how organisms deal with stress and shed light on new therapeutic targets. ■

Received 17 August; accepted 11 November 2015.

- Di Prisco, G. V. *et al.* Translational control of mGluR-dependent long-term depression and object-place learning by eIF2 α . *Nature Neurosci.* **17**, 1073–1082 (2014).
- Han, J. *et al.* ER-stress-induced transcriptional regulation increases protein synthesis leading to cell death. *Nature Cell Biol.* **15**, 481–490 (2013). **This study identified that CHOP and ATF4 form heterodimers, which leads to increased protein synthesis, oxidative stress and cell death.**
- Harding, H. P. *et al.* Ppp1r15 gene knockout reveals an essential role for translation initiation factor 2 α (eIF2 α) dephosphorylation in mammalian development. *Proc. Natl Acad. Sci. USA* **106**, 1832–1837 (2009).
- Hollien, J. & Weissman, J. S. Decay of endoplasmic reticulum-localized mRNAs during the unfolded protein response. *Science* **313**, 104–107 (2006).
- Rutkowski, D. T. *et al.* Adaptation to ER stress is mediated by differential stabilities of pro-survival and pro-apoptotic mRNAs and proteins. *PLoS Biol.* **4**, e374 (2006).
- Lin, J. H. *et al.* IRE1 signaling affects cell fate during the unfolded protein response. *Science* **318**, 944–949 (2007).
- Marciniak, S. J. *et al.* CHOP induces death by promoting protein synthesis and oxidation in the stressed endoplasmic reticulum. *Genes Dev.* **18**, 3066–3077 (2004).
- Malhotra, J. D. *et al.* Antioxidants reduce endoplasmic reticulum stress and improve protein secretion. *Proc. Natl Acad. Sci. USA* **105**, 18525–18530 (2008).
- Song, B., Scheuner, D., Ron, D., Pennathur, S. & Kaufman, R. J. *Chop* deletion reduces oxidative stress, improves β cell function, and promotes cell survival in multiple mouse models of diabetes. *J. Clin. Invest.* **118**, 3378–3389 (2008).
- Zinszner, H. *et al.* CHOP is implicated in programmed cell death in response to impaired function of the endoplasmic reticulum. *Genes Dev.* **12**, 982–995 (1998).
- Kaufman, R. J. & Malhotra, J. D. Calcium trafficking integrates endoplasmic reticulum function with mitochondrial bioenergetics. *Biochim. Biophys. Acta* **1843**, 2233–2239 (2014).
- Babour, A., Bicknell, A. A., Tourtellotte, J. & Niwa, M. A surveillance pathway monitors the fitness of the endoplasmic reticulum to control its inheritance. *Cell* **142**, 256–269 (2010). **This study showed that the function of the ER is regulated by the MAPK Sit2, not UPR pathways, in cell mitosis in budding yeast.**
- Reimold, A. M. *et al.* Plasma cell differentiation requires the transcription factor XBP-1. *Nature* **412**, 300–307 (2001). **This study was the first to demonstrate that XBP1 is required for the differentiation of plasma cells.**
- Zhang, K. *et al.* The unfolded protein response sensor IRE1 α is required at 2 distinct steps in B cell lymphopoiesis. *J. Clin. Invest.* **115**, 268–281 (2005).
- Shaffer, A. L. *et al.* XBP1, downstream of Blimp-1, expands the secretory apparatus and other organelles, and increases protein synthesis in plasma cell differentiation. *Immunity* **21**, 81–93 (2004).
- Hu, C. C., Dougan, S. K., McGehee, A. M., Love, J. C. & Ploegh, H. L. XBP-1 regulates signal transduction, transcription factors and bone marrow colonization in B cells. *EMBO J.* **28**, 1624–1636 (2009).
- Hassler, J. R. *et al.* The IRE1 α /XBP1s pathway is essential for the glucose response and protection of β cells. *PLoS Biol.* **13**, e1002277 (2015).
- Acosta-Alvear, D. *et al.* XBP1 controls diverse cell type- and condition-specific transcriptional regulatory networks. *Mol. Cell* **27**, 53–66 (2007).
- Huh, W. J. *et al.* XBP1 controls maturation of gastric zymogenic cells by induction of MIST1 and expansion of the rough endoplasmic reticulum. *Gastroenterology* **139**, 2038–2049 (2010).
- Lee, A. H., Heidtman, K., Hotamisligil, G. S. & Glimcher, L. H. Dual and opposing roles of the unfolded protein response regulated by IRE1 α and XBP1 in proinsulin processing and insulin secretion. *Proc. Natl Acad. Sci. USA* **108**, 8885–8890 (2011).
- Adolph, T. E. *et al.* Paneth cells as a site of origin for intestinal inflammation. *Nature* **503**, 272–276 (2013).
- Kamimura, D. & Bevan, M. J. Endoplasmic reticulum stress regulator XBP-1 contributes to effector CD8⁺ T cell differentiation during acute infection. *J. Immunol.* **181**, 5433–5441 (2008).
- Iwakoshi, N. N., Pypaert, M. & Glimcher, L. H. The transcription factor XBP-1 is essential for the development and survival of dendritic cells. *J. Exp. Med.* **204**, 2267–2275 (2007).
- Coelho, D. S. *et al.* Xbp1-independent Ire1 signaling is required for photoreceptor differentiation and rhabdomyere morphogenesis in *Drosophila*. *Cell Rep.* **5**, 791–801 (2013).
- Levi-Ferber, M., Gian, H., Dudkevich, R. & Henis-Korenblit, S. Transdifferentiation mediated tumor suppression by the endoplasmic reticulum stress sensor IRE-1 in *C. elegans*. *eLife* **4**, e08005 (2015).
- Wang, X. Z. & Ron, D. Stress-induced phosphorylation and activation of the transcription factor CHOP (GADD153) by p38 MAP kinase. *Science* **272**, 1347–1349 (1996). **This study identified the role of p38 MAPK-mediated CHOP phosphorylation in adipocyte differentiation.**
- Han, J. *et al.* ER stress signalling through eIF2 α and CHOP, but not IRE1 α , attenuates adipogenesis in mice. *Diabetologia* **56**, 911–924 (2013).

28. Bobrovnikova-Marjon, E. *et al.* PERK utilizes intrinsic lipid kinase activity to generate phosphatidic acid, mediate Akt activation, and promote adipocyte differentiation. *Mol. Cell. Biol.* **32**, 2268–2278 (2012).
29. Bobrovnikova-Marjon, E. *et al.* PERK-dependent regulation of lipogenesis during mouse mammary gland development and adipocyte differentiation. *Proc. Natl Acad. Sci. USA* **105**, 16314–16319 (2008).
30. Yu, K. *et al.* Activating transcription factor 4 regulates adipocyte differentiation via altering the coordinate expression of CCAAT/enhancer binding protein β and peroxisome proliferator-activated receptor γ . *FEBS J.* **281**, 2399–2409 (2014).
31. Feng, Y. X. *et al.* Epithelial-to-mesenchymal transition activates PERK-eIF2 α and sensitizes cells to endoplasmic reticulum stress. *Cancer Discov.* **4**, 702–715 (2014).
32. Zhou, A. X. *et al.* C/EBP-homologous protein (CHOP) in vascular smooth muscle cells regulates their proliferation in aortic explants and atherosclerotic lesions. *Circ. Res.* **116**, 1736–1743 (2015).
33. Delépine, M. *et al.* EIF2AK3, encoding translation initiation factor 2- α kinase 3, is mutated in patients with Wolcott-Rallison syndrome. *Nature Genet.* **25**, 406–409 (2000).
34. Harding, H. P. *et al.* Diabetes mellitus and exocrine pancreatic dysfunction in *Perk*^{-/-} mice reveals a role for translational control in secretory cell survival. *Mol. Cell* **7**, 1153–1163 (2001).
35. Scheuner, D. *et al.* Translational control is required for the unfolded protein response and *in vivo* glucose homeostasis. *Mol. Cell* **7**, 1165–1176 (2001).
36. Gao, Y. *et al.* PERK is required in the adult pancreas and is essential for maintenance of glucose homeostasis. *Mol. Cell. Biol.* **32**, 5129–5139 (2012).
37. Back, S. H. *et al.* Translation attenuation through eIF2 α phosphorylation prevents oxidative stress and maintains the differentiated state in β cells. *Cell Metab.* **10**, 13–26 (2009).
38. Usui, M. *et al.* *Atf6a*-null mice are glucose intolerant due to pancreatic β -cell failure on a high-fat diet but partially resistant to diet-induced insulin resistance. *Metabolism* **61**, 1118–1128 (2012).
39. Han, J. *et al.* Antioxidants complement the requirement for protein chaperone function to maintain β -cell function and glucose homeostasis. *Diabetes* **64**, 2892–2904 (2015).
40. Seo, J. *et al.* *Atf4* regulates obesity, glucose homeostasis, and energy expenditure. *Diabetes* **58**, 2565–2573 (2009).
41. Mao, T. *et al.* PKA phosphorylation couples hepatic inositol-requiring enzyme 1 α to glucagon signaling in glucose metabolism. *Proc. Natl Acad. Sci. USA* **108**, 15852–15857 (2011).
42. Birkenfeld, A. L. *et al.* Influence of the hepatic eukaryotic initiation factor 2 α (eIF2 α) endoplasmic reticulum (ER) stress response pathway on insulin-mediated ER stress and hepatic and peripheral glucose metabolism. *J. Biol. Chem.* **286**, 36163–36170 (2011).
43. Li, K. *et al.* MicroRNA-214 suppresses gluconeogenesis by targeting activating transcription factor 4. *J. Biol. Chem.* **290**, 8185–8195 (2015).
44. Wang, Y., Vera, L., Fischer, W. H. & Montminy, M. The CREB coactivator CRTC2 links hepatic ER stress and fasting gluconeogenesis. *Nature* **460**, 534–537 (2009).
45. Wu, J. *et al.* *Atf6a* optimizes long-term endoplasmic reticulum function to protect cells from chronic stress. *Dev. Cell* **13**, 351–364 (2007).
46. Kohl, S. *et al.* Mutations in the unfolded protein response regulator *Atf6* cause the cone dysfunction disorder achromatopsia. *Nature Genet.* **47**, 757–765 (2015).
47. Fu, S. *et al.* Aberrant lipid metabolism disrupts calcium homeostasis causing liver endoplasmic reticulum stress in obesity. *Nature* **473**, 528–531 (2011).
This study established a connection between abnormal lipid and calcium metabolism and hepatic ER stress in obesity.
48. Volmer, R., van der Ploeg, K. & Ron, D. Membrane lipid saturation activates endoplasmic reticulum unfolded protein response transducers through their transmembrane domains. *Proc. Natl Acad. Sci. USA* **110**, 4628–4633 (2013).
49. Lee, J. S., Mendez, R., Heng, H. H., Yang, Z. Q. & Zhang, K. Pharmacological ER stress promotes hepatic lipogenesis and lipid droplet formation. *Am. J. Transl. Res.* **4**, 102–113 (2012).
50. Williams, K. J. *et al.* An essential requirement for the SCAP/SREBP signaling axis to protect cancer cells from lipotoxicity. *Cancer Res.* **73**, 2850–2862 (2013).
51. Rutkowski, D. T. *et al.* UPR pathways combine to prevent hepatic steatosis caused by ER stress-mediated suppression of transcriptional master regulators. *Dev. Cell* **15**, 829–840 (2008).
52. Pottekat, A. *et al.* Insulin biosynthetic interaction network component, TMEM24, facilitates insulin reserve pool release. *Cell Rep.* **4**, 921–930 (2013).
53. Wang, R. *et al.* Insulin secretion and Ca²⁺ dynamics in β -cells are regulated by PERK (EIF2AK3) in concert with calcineurin. *J. Biol. Chem.* **288**, 33824–33836 (2013).
54. Cao, S. S. *et al.* Phosphorylation of eIF2 α is dispensable for differentiation but required at a posttranscriptional level for Paneth cell function and intestinal homeostasis in mice. *Inflamm. Bowel Dis.* **20**, 712–722 (2014).
55. Martinon, F., Chen, X., Lee, A. H. & Glimcher, L. H. TLR activation of the transcription factor XBP1 regulates innate immune responses in macrophages. *Nature Immunol.* **11**, 411–418 (2010).
This study demonstrated that TLR2 and TLR4 activate the IRE1 α -XBP1 pathway to augment pro-inflammatory cytokine production in macrophages.
56. Yamamoto, K. *et al.* Transcriptional induction of mammalian ER quality control proteins is mediated by single or combined action of ATF6 α and XBP1. *Dev. Cell* **13**, 365–376 (2007).
57. Hetz, C. The unfolded protein response: controlling cell fate decisions under ER stress and beyond. *Nature Rev. Mol. Cell Biol.* **13**, 89–102 (2012).
58. Biason-Lauber, A., Lang-Muritano, M., Vaccaro, T. & Schoenle, E. J. Loss of kinase activity in a patient with Wolcott-Rallison syndrome caused by a novel mutation in the *EIF2AK3* gene. *Diabetes* **51**, 2301–2305 (2002).
59. Zhao, L., Longo-Guess, C., Harris, B. S., Lee, J. W. & Ackerman, S. L. Protein accumulation and neurodegeneration in the wozzy mutant mouse is caused by disruption of SIL1, a cochaperone of BiP. *Nature Genet.* **37**, 974–979 (2005).
60. Schroder, M. & Kaufman, R. J. The mammalian unfolded protein response. *Annu. Rev. Biochem.* **74**, 739–789 (2005).
61. Nichols, W. C. *et al.* Mutations in the ER-Golgi intermediate compartment protein ERGIC-53 cause combined deficiency of coagulation factors V and VIII. *Cell* **93**, 61–70 (1998).
62. Hetz, C. & Mollereau, B. Disturbance of endoplasmic reticulum proteostasis in neurodegenerative diseases. *Nature Rev. Neurosci.* **15**, 233–249 (2014).
63. Kondo, T. *et al.* Modeling Alzheimer's disease with iPSCs reveals stress phenotypes associated with intracellular A β and differential drug responsiveness. *Cell Stem Cell* **12**, 487–496 (2013).
64. Baleriola, J. *et al.* Axonally synthesized ATF4 transmits a neurodegenerative signal across brain regions. *Cell* **158**, 1159–1172 (2014).
65. Uehara, T. *et al.* S-nitrosylated protein-disulphide isomerase links protein misfolding to neurodegeneration. *Nature* **441**, 513–517 (2006).
66. Yang, L. *et al.* S-nitrosylation links obesity-associated inflammation to endoplasmic reticulum dysfunction. *Science* **349**, 500–506 (2015).
67. Chung, C. Y. *et al.* Identification and rescue of α -synuclein toxicity in Parkinson patient-derived neurons. *Science* **342**, 983–987 (2013).
68. Credle, J. J. *et al.* α -Synuclein-mediated inhibition of ATF6 processing into COP1 vesicles disrupts UPR signaling in Parkinson's disease. *Neurobiol. Dis.* **76**, 112–125 (2015).
69. Shimura, H. *et al.* Ubiquitination of a new form of α -synuclein by parkin from human brain: implications for Parkinson's disease. *Science* **293**, 263–269 (2001).
70. Vidal, R. L. *et al.* Targeting the UPR transcription factor XBP1 protects against Huntington's disease through the regulation of FoxO1 and autophagy. *Hum. Mol. Genet.* **21**, 2245–2262 (2012).
71. Ozcan, U. *et al.* Endoplasmic reticulum stress links obesity, insulin action, and type 2 diabetes. *Science* **306**, 457–461 (2004).
This study revealed that ER stress is a central feature of insulin resistance, obesity and type 2 diabetes.
72. Wang, S. *et al.* IRE1 α -XBP1s induces PDI expression to increase MTP activity for hepatic VLDL assembly and lipid homeostasis. *Cell Metab.* **16**, 473–486 (2012).
73. Zhang, K. *et al.* Endoplasmic reticulum stress activates cleavage of CREBH to induce a systemic inflammatory response. *Cell* **124**, 587–599 (2006).
This study was the first to provide a link between ER stress and the systemic inflammatory response through CREB-H cleavage to induce acute-phase response genes and pro-inflammatory cytokines in hepatocytes.
74. Lee, A. H., Scapa, E. F., Cohen, D. E. & Glimcher, L. H. Regulation of hepatic lipogenesis by the transcription factor XBP1. *Science* **320**, 1492–1496 (2008).
75. So, J. S. *et al.* Silencing of lipid metabolism genes through IRE1 α -mediated mRNA decay lowers plasma lipids in mice. *Cell Metab.* **16**, 487–499 (2012).
76. Zhang, K. *et al.* The unfolded protein response transducer IRE1 α prevents ER stress-induced hepatic steatosis. *EMBO J.* **30**, 1357–1375 (2011).
77. Ozcan, L. *et al.* Endoplasmic reticulum stress plays a central role in development of leptin resistance. *Cell Metab.* **9**, 35–51 (2009).
78. Huang, C. J. *et al.* High expression rates of human islet amyloid polypeptide induce endoplasmic reticulum stress mediated β -cell apoptosis, a characteristic of humans with type 2 but not type 1 diabetes. *Diabetes* **56**, 2016–2027 (2007).
79. Tersey, S. A. *et al.* Islet β -cell endoplasmic reticulum stress precedes the onset of type 1 diabetes in the nonobese diabetic mouse model. *Diabetes* **61**, 818–827 (2012).
80. Marfouf, I. *et al.* Expression of endoplasmic reticulum stress markers in the islets of patients with type 1 diabetes. *Diabetologia* **55**, 2417–2420 (2012).
81. Lee, J. H. *et al.* The transcription factor cyclic AMP-responsive element-binding protein H regulates triglyceride metabolism. *Nature Med.* **17**, 812–815 (2011).
82. Wang, M. & Kaufman, R. J. The impact of the endoplasmic reticulum protein-folding environment on cancer development. *Nature Rev. Cancer* **14**, 581–597 (2014).
83. Qiu, Q. *et al.* Toll-like receptor-mediated IRE1 α activation as a therapeutic target for inflammatory arthritis. *EMBO J.* **32**, 2477–2490 (2013).
84. Hasnain, S. Z. *et al.* Glycemic control in diabetes is restored by therapeutic manipulation of cytokines that regulate beta cell stress. *Nature Med.* **20**, 1417–1426 (2014).
85. Zhang, K. & Kaufman, R. J. From endoplasmic-reticulum stress to the inflammatory response. *Nature* **454**, 455–462 (2008).
86. Menu, P. *et al.* ER stress activates the NLRP3 inflammasome via an UPR-independent pathway. *Cell Death Dis.* **3**, e261 (2012).
87. Lerner, A. G. *et al.* IRE1 α induces thioredoxin-interacting protein to activate the NLRP3 inflammasome and promote programmed cell death under irremediable ER stress. *Cell Metab.* **16**, 250–264 (2012).
88. Bronner, D. N. *et al.* Endoplasmic reticulum stress activates the inflammasome via NLRP3- and caspase-2-driven mitochondrial damage. *Immunity* **43**, 451–462 (2015).
89. Cao, S. S. *et al.* The unfolded protein response and chemical chaperones reduce protein misfolding and colitis in mice. *Gastroenterology* **144**, 989–1000 (2013).
90. Nakagawa, H. *et al.* ER stress cooperates with hypernutrition to trigger TNF-dependent spontaneous HCC development. *Cancer Cell* **26**, 331–343 (2014).
This study showed the oncogenic role of ER stress in hepatocellular carcinoma tumorigenesis.

91. Engin, F. *et al.* Restoration of the unfolded protein response in pancreatic β cells protects mice against type 1 diabetes. *Sci. Transl. Med.* **5**, 211ra156 (2013).
92. Kars, M. *et al.* Tauroursodeoxycholic acid may improve liver and muscle but not adipose tissue insulin sensitivity in obese men and women. *Diabetes* **59**, 1899–1905 (2010).
93. Xiao, C., Giacca, A. & Lewis, G. F. Sodium phenylbutyrate, a drug with known capacity to reduce endoplasmic reticulum stress, partially alleviates lipid-induced insulin resistance and β -cell dysfunction in humans. *Diabetes* **60**, 918–924 (2011).
94. Han, K. L. *et al.* Therapeutic potential of peroxisome proliferators-activated receptor- α/γ dual agonist with alleviation of endoplasmic reticulum stress for the treatment of diabetes. *Diabetes* **57**, 737–745 (2008).
95. Yusta, B. *et al.* GLP-1 receptor activation improves β cell function and survival following induction of endoplasmic reticulum stress. *Cell Metab.* **4**, 391–406 (2006).
96. Sawkar, A. R. *et al.* Chemical chaperones increase the cellular activity of N370S β -glucosidase: a therapeutic strategy for Gaucher disease. *Proc. Natl Acad. Sci. USA* **99**, 15428–15433 (2002).
97. Zuleta, A., Vidal, R. L., Armentano, D., Parsons, G. & Hetz, C. AAV-mediated delivery of the transcription factor XBP1s into the striatum reduces mutant Huntingtin aggregation in a mouse model of Huntington's disease. *Biochem. Biophys. Res. Commun.* **420**, 558–563 (2012).
98. Flaherty, D. P. *et al.* Discovery of sulfonamidebenzamides as selective apoptotic CHOP pathway activators of the unfolded protein response. *ACS Med. Chem. Lett.* **5**, 1278–1283 (2014).
99. Mahadevan, N. R. *et al.* Transmission of endoplasmic reticulum stress and pro-inflammation from tumor cells to myeloid cells. *Proc. Natl Acad. Sci. USA* **108**, 6561–6566 (2011).
100. Taylor, R. C. & Dillin, A. XBP-1 is a cell-nonautonomous regulator of stress resistance and longevity. *Cell* **153**, 1435–1447 (2013).
This study showed that forced activation of the UPR in neurons could be transmitted to intestine tissues to increase longevity.
101. Williams, K. W. *et al.* Xbp1s in Pomc neurons connects ER stress with energy balance and glucose homeostasis. *Cell Metab.* **20**, 471–482 (2014).
102. Gardner, B. M. & Walter, P. Unfolded proteins are Ire1-activating ligands that directly induce the unfolded protein response. *Science* **333**, 1891–1894 (2011).
103. Carrara, M., Prischi, F., Nowak, P. R., Kopp, M. C. & Ali, M. M. Noncanonical binding of BiP ATPase domain to Ire1 and Perk is dissociated by unfolded protein $C_{\alpha}1$ to initiate ER stress signaling. *eLife* **4**, e03522 (2015).
104. Zhou, J. *et al.* The crystal structure of human IRE1 luminal domain reveals a conserved dimerization interface required for activation of the unfolded protein response. *Proc. Natl Acad. Sci. USA* **103**, 14343–14348 (2006).
105. Fu, S. *et al.* Phenotypic assays identify azoramide as a small-molecule modulator of the unfolded protein response with antidiabetic activity. *Sci. Transl. Med.* **7**, 292ra98 (2015).
106. Kudo, T. *et al.* A molecular chaperone inducer protects neurons from ER stress. *Cell Death Differ.* **15**, 364–375 (2008).
107. Das, I. *et al.* Preventing proteostasis diseases by selective inhibition of a phosphatase regulatory subunit. *Science* **348**, 239–242 (2015).
108. Sekine, Y. *et al.* Mutations in a translation initiation factor identify the target of a memory-enhancing compound. *Science* **348**, 1027–1030 (2015).
109. Sidrauski, C. *et al.* Pharmacological brake-release of mRNA translation enhances cognitive memory. *eLife* **2**, e00498 (2013).
110. Ghosh, R. *et al.* Allosteric inhibition of the IRE1 α RNase preserves cell viability and function during endoplasmic reticulum stress. *Cell* **158**, 534–548 (2014).
111. Tam, A. B., Koong, A. C. & Niwa, M. Ire1 has distinct catalytic mechanisms for XBP1/HAC1 splicing and RIDD. *Cell Rep.* **9**, 850–858 (2014).
112. Bright, M. D., Itzhak, D. N., Wardell, C. P., Morgan, G. J. & Davies, F. E. Cleavage of BLOC1S1 mRNA by IRE1 is sequence specific, temporally separate from XBP1 splicing, and dispensable for cell viability under acute endoplasmic reticulum stress. *Mol. Cell Biol.* **35**, 2186–2202 (2015).
113. Upton, J. P. *et al.* IRE1 α cleaves select microRNAs during ER stress to derepress translation of proapoptotic Caspase-2. *Science* **338**, 818–822 (2012).
114. Sandow, J. J. *et al.* ER stress does not cause upregulation and activation of caspase-2 to initiate apoptosis. *Cell Death Differ.* **21**, 475–480 (2014).
115. Lu, M. *et al.* Opposing unfolded-protein-response signals converge on death receptor 5 to control apoptosis. *Science* **345**, 98–101 (2014).
116. Hur, K. Y. *et al.* IRE1 α activation protects mice against acetaminophen-induced hepatotoxicity. *J. Exp. Med.* **209**, 307–318 (2012).
117. Kaser, A. *et al.* XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. *Cell* **134**, 743–756 (2008).
This study was the first to demonstrate that ER stress in intestinal epithelial cells leads to the development of inflammatory bowel disease.
118. Kamata, H. *et al.* Reactive oxygen species promote TNF α -induced death and sustained JNK activation by inhibiting MAP kinase phosphatases. *Cell* **120**, 649–661 (2005).
119. Cullinan, S. B. *et al.* Nrf2 is a direct PERK substrate and effector of PERK-dependent cell survival. *Mol. Cell Biol.* **23**, 7198–7209 (2003).
120. Lu, P. D. *et al.* Cytoprotection by pre-emptive conditional phosphorylation of translation initiation factor 2. *EMBO J.* **23**, 169–179 (2004).
121. Qiu, Y. *et al.* A crucial role for RACK1 in the regulation of glucose-stimulated IRE1 α activation in pancreatic β cells. *Sci. Signal.* **3**, ra7 (2010).
122. Moreno, J. A. *et al.* Sustained translational repression by eIF2 α -P mediates prion neurodegeneration. *Nature* **485**, 507–511 (2012).
123. Moreno, J. A. *et al.* Oral treatment targeting the unfolded protein response prevents neurodegeneration and clinical disease in prion-infected mice. *Sci. Transl. Med.* **5**, 206ra138 (2013).
124. Halliday, M. *et al.* Partial restoration of protein synthesis rates by the small molecule ISRIB prevents neurodegeneration without pancreatic toxicity. *Cell Death Dis.* **6**, e1672 (2015).
125. Krishnamoorthy, J. *et al.* Evidence for eIF2 α phosphorylation-independent effects of GSK2656157, a novel catalytic inhibitor of PERK with clinical implications. *Cell Cycle* **13**, 801–806 (2014).

Supplementary Information is linked to the online version of the paper at go.nature.com/jfhkeh

Acknowledgements The authors apologize to those researchers whose references they were unable to acknowledge due to space limitations. R.J.K. is supported by US National Institutes of Health grants R37 DK042394, R01 DK088227, R01 DK103183 and CA128814.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this paper at go.nature.com/jfhkeh. Correspondence should be addressed to R.J.K. (rkaufman@sbpdiscovery.org).

Antibacterial drug discovery in the resistance era

Eric D. Brown^{1,2} & Gerard D. Wright^{1,2}

The looming antibiotic–resistance crisis has penetrated the consciousness of clinicians, researchers, policymakers, politicians and the public at large. The evolution and widespread distribution of antibiotic–resistance elements in bacterial pathogens has made diseases that were once easily treatable deadly again. Unfortunately, accompanying the rise in global resistance is a failure in antibacterial drug discovery. Lessons from the history of antibiotic discovery and fresh understanding of antibiotic action and the cell biology of microorganisms have the potential to deliver twenty-first century medicines that are able to control infection in the resistance era.

We are entering a period in infection control that has not been experienced for more than three-quarters of a century. In this era of ever-increasing resistance (Fig. 1), the management of bacterial infection through the use of safe, cheap and plentiful antibiotics can no longer be taken for granted. The serendipitous discovery of penicillin in 1929 (ref. 1) enabled the effective control of infections caused by Gram-positive pathogens such as *Staphylococcus* and *Streptococcus*, and the isolation of streptomycin in 1943 (ref. 2) facilitated the control of tuberculosis agent *Mycobacterium tuberculosis* — for the first time in history. These discoveries propelled the golden era of antibiotics, in which natural scaffolds and alternative versions of these pioneering drugs were uncovered by mining the specialized metabolism of bacteria and fungi or by the chemical modification of existing scaffolds (Table 1). The results were revolutionary.

Released from the hegemony of infection, clinicians began to transform medicine. Invasive surgeries became routine, immune-system-shattering chemotherapy was introduced to fight the ‘war’ on cancer, organ transplantation extended lives and the radical replacement of deteriorating joints, diseased corneas and burnt skin improved quality of life for millions of people. In the resistance era, these breakthroughs — and the consequent improvements in our quality of life and life expectancy — are at risk.

The microbial world has always had the molecular tools to drive resistance. Surveys of microorganisms in the environment and in samples of ancient permafrost revealed that the antibiotic resistome — the global collection of all microbial resistance genes — is genetically diverse^{3,4}, widespread across all environmental niches^{5–8} and pre-dates the modern antibiotic era by millennia⁹. The use of large quantities of antibiotics to control infection in human and animal diseases and in agriculture has created unprecedented conditions for the mobilization of resistance elements in bacterial populations and their capture by previously antibiotic-sensitive pathogens. Over time, these conditions have enabled the selection of evermore drug-resistant bacteria through three main mechanisms: the sequential capture of a myriad of resistance genes, mostly through mobilization and horizontal transfer from environmental sources; the ‘freezing’ of polymorphisms in antibiotic target genes through secondary mutations that neutralize the effects on fitness that drug-resistance mutations often impart¹⁰; and the upregulation of intrinsic mechanisms of resistance, such as efflux^{11,12} or antibiotic-inactivating enzymes¹³. Multidrug resistance

in pathogens is typical in the resistance era¹⁴ and is eroding our ability to control infections with antibiotics. Pathogenic strains of bacteria that are resistant to most, or all, available antibiotics are now isolated routinely. We have entered the post-antibiotic age.

Mechanisms and origins of antibiotics

Antibiotics perturb important biochemical processes, which results in the inhibition of cell growth and division and, in the case of bactericidal agents, cell death¹⁵. The first antimicrobial agents were synthetic molecules that were discovered by screening libraries of chemicals, in particular dyes. This screening was superseded by the realization that bacteria and fungi in the environment produce metabolites that could treat bacterial infections in humans with remarkable efficacy and minimal toxic side effects. The strategy adopted by Selman Waksman, a co-discoverer of streptomycin, involved screening soil-dwelling bacteria, and in particular the spore-forming Actinomycetes, for the production of metabolites that block the growth of pathogens — a process that was later termed the Waksman platform¹⁶. The simplicity and effectiveness of the platform ushered in the golden era of antibiotic discovery, the period in which most of the microbial natural scaffolds that serve as our antibiotic arsenal were discovered (Fig. 1). The platform also imposed measures of success that subsequent drug-discovery efforts would need to use, in particular use of the inhibition of cell growth *in vitro*, which is assessed on rich media, as the main way to determine the minimal inhibitory concentration (MIC) of a compound.

By the mid-1960s, new and effective antibiotic scaffolds were becoming harder to identify using the Waksman platform. Because these specialized metabolites are the product of microbial evolution within a specific environment and were not designed as drugs, most had considerable pharmacological or toxicological drawbacks. And resistance to these early antibiotics, the result of the horizontal transfer of resistance genes between bacteria or chromosomal mutation, was also becoming a problem. These issues spawned the era of medicinal chemistry, the next period of innovation in antibiotic discovery (Fig. 1). The development of antibiotics in this period was dominated by cycles of innovation that focused largely on creating synthetic versions of the natural scaffolds of the golden era of discovery. These derivatives led to outstanding improvements in the application of antibiotics, which included lower doses, an expanded antimicrobial spectrum against various pathogens, and the avoidance of resistance.

¹Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Ontario L8S 4L8, Canada. ²Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario L8S 4L8, Canada.

These newer drugs substantially changed the practice of medicine and ushered in the ‘miracles’ of modern medicine that we now take for granted.

Most existing antibiotics are therefore derived from natural products and tend to target the bacterial cell wall, DNA or ribosomes. With a few exceptions, these compounds exert pleiotropic and complex effects on the bacterial cell and often have more than one molecular target. The β -lactam antibiotics, such as penicillin, covalently modify a number of target enzymes known as penicillin-binding proteins (PBPs). Collectively, these enzymes are responsible for the synthesis and remodelling of the bacterial cell wall for growth and division. Antibiotics that inhibit protein synthesis target the ribosome, which is encoded by multiple copies of ribosomal DNA, and those that block DNA synthesis act on several topoisomerase enzymes. Importantly, this effect on multiple cellular targets limits the frequency of spontaneous resistance that can arise from mutation in the target gene. Inhibition of the molecular targets of antibiotics commonly results in complex downstream effects that exceed those of simple enzyme inhibition. Mounting evidence suggests that β -lactam antibiotics disorganize the activity of the bacterial cell-wall synthesis machinery in a manner far more complicated than simple inhibition¹⁷. The ribosome is a macromolecular complex with many enzymatic functions, regulatory sites and components. When it is targeted by aminoglycosides, for example, the resulting synthesis of aberrant proteins leads to pleiotropic and toxic effects on the bacterial cell^{18,19}. Systems-biology approaches suggest that reactive oxygen species have been overlooked as contributors to cell death²⁰. Although this hypothesis remains controversial, there is an increasing appreciation that bacterial cell death is complex and probably requires the involvement of several cellular pathways. Many natural-product antibiotics are the product of selection for these complex traits over millions of years of evolution. Therefore, it is perhaps unsurprising that modern methods of drug discovery have yet to deliver compounds with efficacy comparable to that of the first generation of natural antibiotics and their semisynthetic derivatives.

Modern antibacterial drug discovery

Successes in the golden era were followed by a considerable hiatus in the identification of new scaffolds that began in the 1960s and lasted until the early 1990s, when a renaissance in discovery efforts was precipitated by a fresh wave of resistance. This coincided with the emergence of innovative drug-discovery approaches in all therapeutic areas. These new methods were underpinned by remarkable advances in technology, such as breakthroughs in the manipulation of recombinant DNA to produce desired proteins at high yields and in high-throughput synthesis to create large chemical libraries. Improvements in facile protein-structure determination enabled rational drug design, and robotic liquid handling facilitated high-throughput screens of biochemical assays. Furthermore, a computing revolution made it possible to handle much larger data sets. When coupled with emergent genomic technologies, these advances enabled a ‘genes-to-drugs’ model that has been applied with success to many therapeutic areas²¹ and become the convention. Yet more than two decades into the practice of modern antibacterial drug discovery, no new medicines have been discovered through this approach.

Antibacterial drug discovery has been especially influenced by genomic data, technology and innovation. The first free-living organism to be sequenced was the pathogen *Haemophilus influenzae*²² and, since then, thousands of bacterial genomes have been sequenced. High-throughput techniques that can create precise deletions at the genome scale have been used to explore methodically the dispensability of each gene in both the model Gram-positive bacterium *Bacillus subtilis*²³ and in the model Gram-negative bacterium *Escherichia coli*²⁴. Pathogens such as *Staphylococcus aureus* and *Pseudomonas aeruginosa* have also been the subject of systematic approaches to mutagenesis^{25,26}, as is the fastidious pathogen *M. tuberculosis*²⁷. The goal of these large-scale surveys has been to identify new targets for a

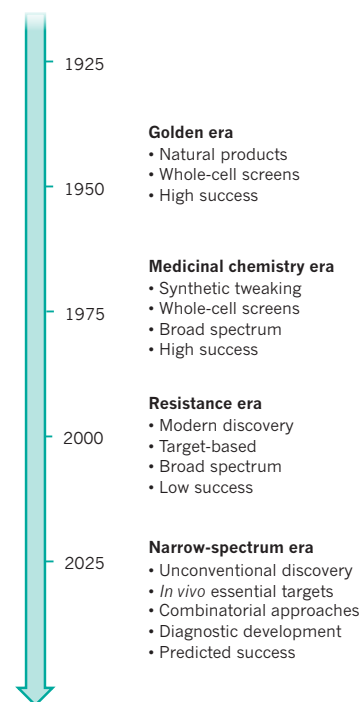


Figure 1 | Models of antibiotic drug discovery and development. Organized with respect to the prevalent models of each era, this timeline highlights the history and future of antibiotic drug discovery. The golden era was initiated by the discovery of sulfonamide and penicillin and is typified by the methods of Selman Waksman and his contemporaries, who used whole-cell screening of natural product extracts to find new antibiotic scaffolds. In subsequent years, termed the medicinal-chemistry era, these scaffolds were modified chemically. Emphasis in the resistance era has fallen on target-based drug discovery to find broad-spectrum agents — a model that has failed to provide new antibiotics. In the future, a focus on innovative methods and unconventional targets should help to create narrow-spectrum agents and associated diagnostics.

generation of antibiotics that are insusceptible to existing mechanisms of resistance²⁸.

In the face of increasing resistance, the failure to yield new antibiotics has been the cause of much concern. Unsuccessful antibacterial programmes have been documented by industry experts^{29–31} and common themes are evident. Large-scale gene dispensability studies generated numerous targets — 160 in one account²⁹ — and many were tested in *in vitro* screening campaigns. However, these high-throughput biochemical screens of large collections of synthetic chemicals often were unable to find promising compounds with the necessary physical and chemical properties. When encouraging lead compounds were found, these failed, largely because of difficulties in developing antibacterial drugs with a suitable product profile — for example, efficacy against a broad spectrum of pathogens or against difficult-to-treat Gram-negative organisms. Gram-negative bacteria have proved especially daunting owing to a poor understanding of drug penetration and efflux systems that are pervasive in this group of pathogens. There were also concerns about the rapid evolution of resistance that is often evident in these single-target approaches. Aside from its emphasis on target-based efforts, modern antibacterial drug discovery has also employed large chemical screens of whole cells for growth inhibition. However, such screens produce huge numbers of active compounds that are difficult to follow up because of a lack of tools for prioritization. Indeed, there has been a resurgence in cell-based approaches that includes innovative strategies for focusing on promising leads that target different aspects of bacterial survival³². Ultimately, these approaches will focus on specific targets to facilitate the lead-optimization strategies that are inherent to modern drug discovery. Therefore, it will become increasingly important to understand the failures of target-based approaches.

The essential-gene paradox

With its focus on the optimization of lead compounds that target individual proteins using state-of-the-art tools of chemistry, biochemistry and pharmacology, modern industrial drug discovery is an inherently reductionist endeavour. Thus, an important premise is that the target is valid, and the weight of evidence to support a campaign against any given target can vary considerably. For example, although systematic surveys of gene dispensability in bacteria have illuminated a large number of potential targets, a substantial fraction of these genes encode proteins of unknown function. A clear understanding of function is commonly inherent in the process of target validation and — from a practical standpoint — it is difficult to build a chemical screening assay for a protein of unknown function. We are coming to understand that dispensability phenotypes are contextual even for well-characterized protein targets, and that growth conditions and genomic context can greatly influence the dispensability of the corresponding gene³³.

In vivo essential genes

Approximately 7% of the *E. coli* genome, some 303 genes, has been shown to be essential for growth in rich media, which represents typical Waksman-screen conditions²⁴. Nevertheless, systematic studies of stress conditions such as nutrient deprivation and chemical perturbation have demonstrated that a further 258 genes are conditionally essential^{34,35}. These conditionally essential genes largely encode enzymes that are important for the utilization of carbon sources or the biosynthesis of the outer membrane as well as in the synthesis of amino acids, vitamins and nucleobases. Yet these processes have long been ignored by antibacterial drug discovery. Indeed, comprehensive efforts to define essential sets of bacterial genes have been performed only *in vitro* and, most extensively, using model microorganisms. Thus, the set of genes that is required for the viability of pathogens in the context of infection is less well understood. These *in vivo* essential genes have been assessed in various pathogens using random transposon mutagenesis approaches that, coincidentally, were invented while the first bacterial genomes were being sequenced³⁶. The methodology has revealed candidate roles for hundreds of genes in aspects of virulence or colonization, including the bacterial stress response, nutrient biosynthesis, type III secretion and attachment functions that are mediated by pili, fimbriae or surface carbohydrates³⁷. Typically, these approaches assess large pools of mutants and have been enabled by the introduction of massively parallel DNA sequencing³⁸. An intriguing application of this technology was an investigation of the genomic requirements for growth of *P. aeruginosa* in sputum

from a person with cystic fibrosis³⁹. Vitamin biosynthesis was found to be an important requirement for bacterial viability — noteworthy because searches for antibacterial chemicals have overwhelmingly emphasized rich media conditions.

In vivo essential genes represent a set of emerging targets that remain untested in modern antibacterial drug discovery. However, they hold great promise for expanding the pool of targets for new drugs. Inhibitors of this class of targets often fail to show the standard phenotype of cell-growth inhibition, which is used to calculate the MIC. The MIC is an important driver of compound optimization in preclinical antibiotic development and a gold standard for efficacy that is understood by regulators, discoverers and practitioners alike. In the absence of a MIC, an assay that is amenable to high-throughput screening and downstream drug discovery needs to be built. One innovative solution involves screening in a non-conventional growth media. This approach was used to target the glyoxylate shunt of *P. aeruginosa* because of its involvement in pulmonary infection⁴⁰. The screening workflow prioritized compounds that were active in nutrient-limited media that contained acetate as the only source of carbon but inactive when glucose was the only source, and led to the discovery of eight lead compounds. Another study built on a platform to discover small molecules that block growth solely under nutrient-limited conditions, and used it to target the biosynthesis of amino acids, nucleobases and vitamins⁴¹. After prioritizing antibacterial compounds that were active only in the absence of nutrient supplements, the approach used systematic supplementation with individual and pools of metabolites to illuminate mechanisms of action. Lead compounds that target glycine, folate and biotin synthesis in *E. coli* were uncovered as a result. Research that targets the quorum-sensing virulence pathway in the pathogen *P. aeruginosa* has generated promising compounds that are active in mouse models of infection without perturbing growth *in vitro*⁴². Through an innovative screening platform, the quorum-sensing system and expression of the toxic gene product SacB were linked such that 'hits' in the screen led to cell viability. Another route to the discovery of *in vivo* targets involves screening in host models of disease. For example, a high-content microscopic screen of macrophages that were infected with *M. tuberculosis* identified a series of lead compounds that targeted the cytochrome *bc₁* complex in the bacterium⁴³.

The essential-gene paradox reveals a fascinating interplay between risk and the target set of genes that has consequences for industrial drug discovery (Fig. 2). For instance, a pathogen might have a set of genes that is essential for growth on rich media and includes conventional targets that are already under investigation. A greater innovation risk would be to

Table 1 | Unique chemical classes of antibiotics discovered in the golden era

Chemical class	Target	Mode of action	Examples
Sulfonamides*	Folate synthesis	Bacteriostatic	Sulfanilamide
β-Lactams†	Cell-wall synthesis	Bacteriocidal	Penicillins Cephalosporins Carbapenems
Aminoglycosides†	Protein synthesis	Bacteriocidal	Spectinomycin Kanamycin Neomycin
Tetracyclines†	Protein synthesis	Bacteriostatic	Tetracycline Doxycycline
Chloramphenicol†	Protein synthesis	Bacteriostatic	Chloramphenicol
Macrolides†	Protein synthesis	Bacteriostatic	Erythromycin Clarithromycin
Glycopeptides†	Cell-wall synthesis	Bacteriocidal	Vancomycin Teicoplanin
Oxazolidinones*	Protein synthesis	Bacteriostatic	Linezolid
Ansamycins†	RNA synthesis	Bacteriocidal	Rifamycin
Quinolones†	DNA synthesis	Bacteriocidal	Ciprofloxacin
Streptogramins†	Protein synthesis	Bacteriocidal	Pristinamycin

*Synthetic chemical †Natural product.

place the focus on the discovery of targets that are required for growth on nutrient-limited media. Such targets are *in vivo* essential genes but would also have *in vitro* phenotypes that facilitate conventional approaches to antibiotic drug development. And an even greater innovation risk would be to focus on genes that are uniquely essential for infection. Studies so far suggest that infection genes could represent the largest set of targets. However, the lack of *in vitro* phenotypes, such as growth inhibition, means that it can be a challenge to enable established methods of drug discovery for such targets.

Genetic–interaction networks

As modern drug discovery has come of age, we have witnessed a transformation in our understanding of the cell that is changing our view of the druggable genome. Studies of gene and protein interactions in yeast^{44,45} and bacteria^{46,47} have revealed a network of abundant functional interactions. The classic maps of biochemical metabolism and cell signalling are being supplanted by a network model of the cell that is characterized by a highly connected web of proteins and genes that exhibits remarkable complexity and redundancy. In the model micro-organism *S. cerevisiae*, in which genome-scale experiments to create a matrix of pairwise deletions for all 5,000 dispensable genes are nearly complete, around 200,000 pairs of genes were revealed to be synthetic lethal *in vitro*⁴⁴. These double-deletion strains suffered a substantial deficiency in fitness compared with the corresponding single-deletion strains, for which the network revealed about 40 interactions per gene. This is in contrast with the approximately 1,000 essential genes that have been mapped for this yeast. Early studies of synthetic lethality in the model bacterium *E. coli* revealed a similar density of gene interactions⁴⁷ and predicted a dramatically expanded body of targets for drugs or combinations of drugs that inhibit interacting gene products. Indeed, these realities make a compelling case for combinatorial approaches to new antibiotics.

The contextual nature of gene essentiality is perhaps best demonstrated by synthetic-viable gene interactions. Here, the double mutant is considerably more fit than expected with respect to the fitness of the corresponding single mutants. An example is the interaction of toxin and antitoxin genes that have a role in bacterial stress responses⁴⁸. The gene that encodes the antitoxin has an essential phenotype, but becomes dispensable in a genetic background that lacks the toxin gene. Another, more remarkable, example of synthetic viability can be found in the wall teichoic acid (WTA) biosynthetic pathway of Gram-positive bacteria. Although early steps in the pathway are dispensable, genes that encode late-step enzymes have an essential phenotype⁴⁹. Curiously, the deletion of an early gene renders any of the late genes dispensable. The essential phenotype of the late genes is thought to arise from the accumulation of WTA biosynthetic intermediates that are linked to undecaprenyl moieties and the sequestration of undecaprenyl phosphate from its role as a lipid carrier in the truly essential process of peptidoglycan-wall assembly⁵⁰. The unique dispensability patterns of WTA genes have provided the basis for sophisticated platforms for the discovery of new lead compounds that target early and late steps in the WTA pathway in *S. aureus*^{51–53} as well as the related pathway of undecaprenyl-phosphate synthesis⁵⁴. Primary and secondary screens for antagonism that are enabled by synthetic-viable phenotypes facilitate the rapid elimination of nuisance compounds that often affect whole-cell screening campaigns. By taking advantage of such knowledge, innovative screens for new antibiotics can be developed.

Chemical genomics

Further examples of the complexity of bacterial viability are provided by studies that map enhancers or suppressors of antibiotic action on a genome scale^{55–57}. Such efforts have revealed that antibiotics have dense and complicated networks of chemical and genetic interactions that seem to be designed to resist perturbation (Fig. 3). These networks offer a rich vista of new targets for antibiotic adjuvants, which are nonantibiotic compounds that enhance the activity of antibiotics or help to overcome

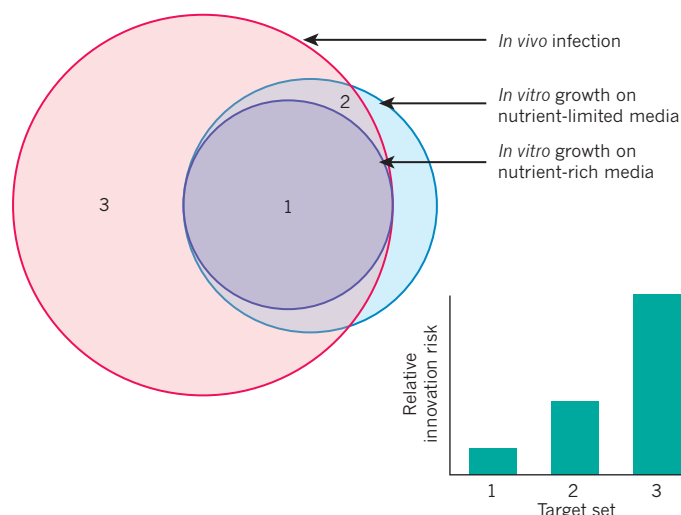


Figure 2 | Target gene sets and innovation risks for a bacterial pathogen.

A Venn diagram to show a relative profile of target genes for a hypothetical bacterial pathogen. Three potential target sets can be imagined. Each has a relative innovation risk within the context of modern industrial drug discovery, which is indicated on the graph. By far the biggest set can be found in the genes that are needed for infection in a relevant animal model, but much validation remains to be done for these targets. The set that is required for growth on nutrient-limited media is well charted. However, not all of the genes in this set will be required for *in vivo* infection. Antibiotic drug discovery under nutrient-rich conditions has been the convention for many years, and the corresponding set of target genes is the smallest.

resistance. Therefore, emerging knowledge of bacterial cell systems is challenging the modern genes-to-drugs approach through the idea that target-validation measures must capture the implications for the network as a whole. An important example of this is the growing appreciation for the role of metabolic states and reactive oxygen species in the effectiveness of existing antibiotics^{58,59}. Indeed, the inherent complexity of cell systems is a compelling argument for the use of whole-cell, phenotype-based screens over biochemical, target-focused efforts in the discovery of antibacterial compounds. Once a new potential compound has been identified, chemical-genomic characterization can provide important information and predictions about its mechanism and resistance profile. Compounds that are discovered and characterized in this way therefore provide a ready probe for target validation and a greater understanding of the network that underpins the target.

Combinatorial discovery

The effectiveness of combinations of antibiotics has been demonstrated in the clinic and is gaining favour as a design strategy. An effective pairing is trimethoprim and sulfamethoxazole, which have a spectacularly synergistic impact on folate metabolism. Another combination combines amoxicillin, a β -lactam antibiotic, with clavulanic acid, an inhibitor of β -lactamase drug resistance enzymes. Clinicians have long combined antibiotics from different classes, such as β -lactams with aminoglycosides, to achieve synergy, to cover a broad spectrum of pathogens when the infectious agent is unknown or to suppress the emergence of resistance. In fact, combinations are the mainstay of treatment for stubborn infections such as tuberculosis. With the exception of β -lactam– β -lactamase inhibitor pairs, existing antimicrobial drug combinations are largely the product of clinical experimentation with established antibiotics. However, a new model of discovery is emerging that recognizes the network view of the cell and searches for efficacious combinations of small molecules in a systematic manner.

Platforms that aim to discover active combinations have almost exclusively searched for adjuvants. This is because the systematic examination of combinations, even in relatively small chemical collections, leads to prohibitively large screens. Typically, an existing antibiotic is tested for

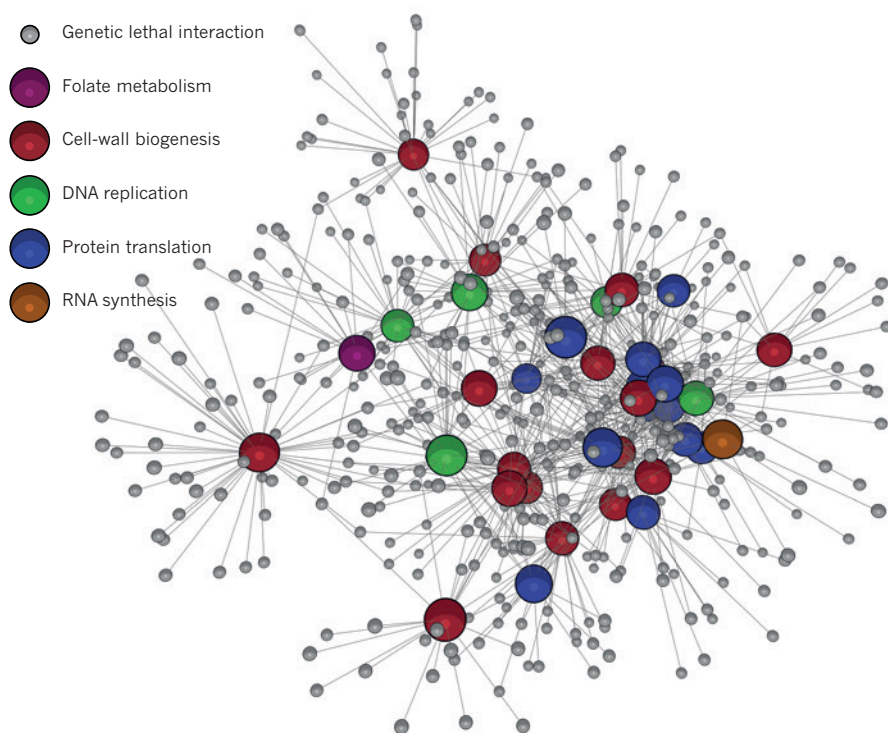


Figure 3 | A chemical–genomic interaction network. The network map illustrates the dense connectivity of enhancing interactions that affect the activity of 39 antibiotics that target the bacterium *Escherichia coli*. Grey balls represent genes and coloured ones are for antibiotics of various mechanistic classes. Interactions are denoted by connecting lines. The map is derived from data in ref. 35 and was prepared with the BioLayout Express^{3D} network visualization and analysis tool⁸³.

growth inhibition at subinhibitory concentrations in combination with a chemical collection. A screen for synthetic compounds that augment the activity of the narrow-spectrum Gram-positive antibiotic novobiocin in the Gram-negative bacterium *E. coli* identified four new compounds⁶⁰. They were found to affect cell shape and membrane permeability, which provided an explanation for the observed synergies. Another combination screen that used natural product antibiotics aimed to reverse carbapenem resistance in metallo- β -lactamase-positive Gram-negative pathogens⁶¹. It found that aspergillomarasmine A could block the notorious carbapenemase NDM-1 and was efficacious in a mouse model of infection by NDM-1-containing *Klebsiella pneumoniae*.

Drugs that have already been approved have attracted considerable interest as sources of bioactive chemical matter for combination studies⁶². This stems from a growing understanding that small molecules with proven therapeutic activity are privileged chemicals that bind to biological polymers and often have potential for alternative uses. Compounds that are used in the clinic have a proven track record in patients and are accompanied by a deep record of study that offers a substantial edge over chemicals for which little is known. Several discovery campaigns that used such libraries have yielded exciting combinations of drugs. For example, the antiarrhythmic drug loperamide was shown to destabilize the membrane potential of bacterial cells, which facilitated the uptake of the antibiotic minocycline in Gram-negative bacteria⁶³. And the antihypertensive drug ticlopidine was found to strengthen the action of β -lactam antibiotics against methicillin-resistant *S. aureus* (MRSA) by inhibiting the synthesis of WTA⁵³.

Combinatorial approaches are consistent with an increased understanding that the most successful antibiotics seem to interact with more than one target³⁰. Furthermore, such approaches are aligned with the network view of bacteria and their resistome, which resides within the cell to buffer against perturbation. Nevertheless, they have received relatively little attention from industry. Indeed, modern target-based discovery approaches seek overwhelmingly to achieve exquisite target selectivity to minimize toxicity that might arise from drug promiscuity.

The adjuvant approach offers advantages such as the potential to decrease the concentration of antibiotic that is needed to achieve a therapeutic effect, which should reduce the emergence of resistance and lessen issues of off-target toxicity. Combinations can extend the effective lifespan

of legacy drugs, and because the delivery and utility of such drugs are familiar to clinicians, barriers to entry into clinical practice might be minimized. However, the combination approach also presents challenges. These include the need to match the pharmacology and efficacy of each component drug, which adds to the burden of drug discovery. Resistance is also a concern — as it is for any antimicrobial therapy. The potential for unexpected drug–drug interactions must also be examined thoroughly. Nevertheless, multicomponent therapies, which are well established in treatment of cancer and of HIV, tuberculosis and bacterial infections — with combinations of β -lactam antibiotics and β -lactamase inhibitors, in particular — offer an alternative to traditional antibiotic discovery in the resistance era.

What chemical matter should be the focus?

The history of antibiotic discovery suggests that natural products, especially those produced by microorganisms, represent privileged chemical matter for the discovery of antibiotics. Because they are the result of natural selection, these compounds have been proved to be highly effective and are the source of most of our antibiotic medicines. Yet they have many drawbacks. Natural-product antibiotics are often chemically complex with challenging and intractable routes to synthesis in the laboratory that make it difficult to prepare derivatives. The pharmacology of most first-generation antibiotics is also not ideal, which reflects the origins of these drugs as specialized microbial metabolites not medicines. Furthermore, because these ancient metabolites are important components of microbial chemical ecology, genetic elements that provide resistance against them have been retained by and are often widely dispersed in communities of microorganisms in the environment^{64,65}. Finally, although screens that use the Waksman platform can readily detect antimicrobial metabolites from libraries of environmental microbes, most such metabolites have already been identified and catalogued. More than 20,000 microbial natural products have been described over the past several decades⁶⁶. Waksman-platform screens of typical libraries of natural products that are derived from the actinomycetes group of bacteria, a favoured source of antibiotics, tend to result in the identification of known compounds⁶⁷. A further complication is that the collection of libraries of natural-product producing microorganisms, the generation of compound extracts for screening and the purification and characterization of active compounds is very labour intensive and costly. Overall, this has resulted in the abandonment

of microbial natural products by most large pharmaceutical companies.

However, libraries of synthetic chemicals can exceed several million compounds and, in principle, provide access to a vast spectrum of chemical space. By definition, they are chemically tractable, which enables their scale up and relatively facile expansion by medicinal chemistry to explore structure–activity relationships, improve efficacy and optimize pharmacology. Indeed, one of the most successful classes of antibiotics over the past four decades has been the synthetic fluoroquinolones. Synthetic libraries are the mainstay of most pharmaceutical companies and are the source of many medicines that are crucial for human health, including inhibitors of protein kinases, G-protein-coupled receptors and ion channels. Although these collections have been explored as sources of new antibiotics using modern drug-discovery platforms for more than three decades, they have yet to deliver new medicines — partly because they are sampling such a narrow area of chemical space. Driven by metrics such as the Lipinski rule of five for orally available drugs⁶⁸, the synthetic libraries of most pharmaceutical companies tend to be dominated by compounds that are optimized for human biology. But most antibiotics in clinical use do not conform to these parameters⁶⁹ because they are natural products or their derivatives. The Achilles' heel of synthetic compounds as new antibiotic drugs seems to be their general inability to penetrate readily the cell envelope of bacteria, which comprises inner and outer membranes, porins and complex carbohydrate polymers, coupled with their susceptibility to active efflux by membrane-associated pumps. Because they are products of evolution, antibiotic natural products have evolved to overcome these challenges. The overall result is a stalemate in discovery: screens of natural-product libraries identify bioactive but known compounds, and screens of synthetic libraries identify potent ligands of biochemical targets but with poor bioactivity.

A solution to this impasse is the development of synthetic libraries that capture the chemical diversity and physicochemical properties of natural products. These are being pursued for a number of therapeutic classes⁷⁰, including antibiotics^{71,72}. Another solution is to capitalize on

the ability of synthetic compounds to inhibit essential bacterial targets by developing delivery systems that solve the cell-envelope penetration and efflux challenges. Conjugation to siderophores is an example of such a 'Trojan horse' approach that has been explored but not yet proved to be clinically efficacious⁷³. A further solution is to combine these compounds with molecules that enhance their transport or otherwise help to breach the permeability barrier of the bacterial cell. In-depth study of the bacterial cell envelope and the development of rules to guide the synthesis of potential antibiotics should enhance the ability of such compounds to penetrate cells and avoid efflux.

The crisis in antibiotic discovery is prompting a return to the investigation of natural products. Known chemical scaffolds that were abandoned in the past are being revisited with some success. Daptomycin, for example, is now being used to treat serious infections caused by Gram-positive pathogens. Eli Lilly and Company had discarded the compound in the 1980s owing to toxic adverse effects that were identified during clinical trials. But in the late 1990s, Cubist Pharmaceuticals licensed the compound and were able to overcome the adverse effects by altering the formulation and dosage. The US Food and Drug Administration (FDA) approved the drug in 2003 (ref. 74), and by 2014, annual sales of daptomycin had exceeded US\$1 billion. Similarly, the poorly absorbed antibiotic fidaxomicin, which was discovered in the 1970s at Hoechst Marion Roussel, saw no clinical use until infections associated with the gut bacterium *Clostridium difficile* started to rise around a decade ago⁷⁵. Fidaxomicin's shortcomings as a broad-spectrum, orally available antibiotic were revisited as strengths in an agent that targets *C. difficile* by taking advantage of high concentrations of the drug in faeces, and fidaxomicin was approved for use by the FDA in 2011.

Over the past 20 years, we have taken amazing strides in our understanding of the mechanisms and genomics of microbial natural-product biosynthesis. Genome sequencing of microorganisms has revealed a spectrum of potential for specialized metabolite production that greatly

BOX 1

Hurdles and solutions for antibiotic discovery

If not for the emergence of resistance to existing antibiotics, these medicines might be thought of as 'perfect' therapeutics with an extraordinary record of success in the clinic. Thus, the bar is set high for the next generation of these medicines. Here are four hurdles and solutions to the antibiotic-discovery problem.

- **Resistance has arisen partly because only a narrow selection of chemical compounds are available, and they have a limited range of mechanisms. Genomics-driven efforts to exploit new potential targets, however, have been unsuccessful.** New targets and approaches to antibiotic drug discovery are fraught with risk. Thus, there is a substantial need for fundamental research to mitigate them. Academic research on unconventional targets and discovery platforms has made much progress in past decades. However, most resources in new antibiotic drug discovery have been invested conservatively in conventional targets and approaches. With growing potential for unconventional targets and new approaches to antibiotic drug discovery, there is great opportunity for successful efforts to industrialize these platforms.
- **The complex mechanisms of action of existing antibiotics could be crucial to their success and elusive to modern drug-discovery platforms.** The solution is to understand these mechanisms better. Biology is complex and our understanding of the action of antibiotics is incomplete. Research that uses both traditional and systems approaches will be required to expand our knowledge and, if this hypothesis proves to be true, our approaches to drug discovery will

need to adapt. In fact, promising new strategies are already being used in the search for combinations of compounds that will target cell networks in ways that are reminiscent of the complex action of existing antibiotics.

- **Physical and chemical properties of existing antibiotics are not compatible with conventional collections of compounds and medicinal-chemistry approaches.** There is great need for research into the antibiotic permeability barriers of bacteria, particularly of Gram-negative organisms, and the physical properties of chemicals that can overcome these barriers. Those who have a role in drug discovery must look critically at existing compound collections and assess medicinal chemistry strategies within this context. Natural sources of chemical compounds have been eschewed in past decades in favour of synthetic chemicals that are better suited to conventional models of drug discovery. It is time to revisit natural products as an alternative to synthetic collections and to embrace new technologies that will help to overcome problems of compound rediscovery.
- **The yardstick for early stage antibiotic discovery is the inhibition of cell growth and the determination of the MIC.** The ease of identifying antibiotics through the Waksman platform, and its success in the golden era, solidified a protocol that narrowly focused success on this simple laboratory test. The realization that numerous targets, such as virulence factors and aspects of metabolism, are not captured by the traditional MIC approach requires innovation in the laboratory and an acceptance that twenty-first century antibiotics might not have the same properties as existing drugs.

exceeds the diversity of compounds that have been purified and characterized over the past century. Automated prediction of the structure and regulation of compounds from genome data is improving rapidly. The manipulation of products with techniques from synthetic biology is becoming more common, and offers an unprecedented opportunity for tapping into the bioactive chemical space. Techniques such as metagenomics and use of specialized growth chambers are enabling exploration of the microbial dark matter — microorganisms that have resisted easy cultivation in the laboratory⁷⁶ — as well as identification of new chemistry⁷⁷ and fresh leads for antibiotic drugs⁷⁸.

Of the thousands of natural-product antibiotics that have been identified over the past decades, only a select few have been seriously investigated as drugs. The examples that we describe in this Review suggest that it would be fruitful to revisit some of these old scaffolds and to search for new ones using twenty-first century technology through the lens of the clinical needs of the resistance era. Large-scale screens will be needed to identify these candidate compounds. At present, no collection of microbial natural products adequately reflects the chemical diversity that has already been reported in the literature, let alone that which is yet to be discovered. Indeed, many 'old' compounds are essentially lost to modern drug discovery because there is no supply of them available. A multi-national effort to develop a truly representative library of antimicrobial natural products that are readily available to antibiotic researchers would be transformative.

The future

Ultimately, clinical need and the marketplace will dictate the product profile of new antibiotics. The clinical need for such drugs is growing, and in some cases, such as multidrug-resistant Gram-negative pathogens, we are already experiencing a profound crisis. In the resistance era, we must accept and tackle this global challenge to continue to enjoy the benefits of antimicrobial agents and the improvements in quality of life and life expectancy that they bring. Although considerable, the regulatory challenges of developing new antibiotics are being slowly addressed to streamline approvals for drugs. For example, the FDA approved a combination treatment that contains ceftazidime and avibactam on the basis of data gathered in phase II clinical trials, therefore avoiding the need for phase III data. And the economic hurdles associated with drugs that cure rather than control diseases are being addressed by proposals such as the GAIN (Generating Antibiotics Incentives Now) act in the United States, which seeks to extend market exclusivity for new antibiotics⁷⁹, public-private partnerships such as the Innovative Medicines Initiative in Europe, and by incentives such as financial rewards for companies that successfully launch new antimicrobial drugs⁸⁰.

The scientific challenges in identifying and developing new antibiotic drugs remain substantial (Box 1). It is clear from the past two decades of effort that antibiotics that are highly effective, safe and broad spectrum are incredibly difficult to find. Consequently, we must ask whether our current medical practices and ideas about infection are appropriate in the resistance era. The availability and properties of the antibiotics of the golden and medicinal-chemistry eras radically changed the practice of infection control and treatment. For example, the efficacy of antibiotics enabled empirical therapy and prophylaxis. And broad-spectrum antibiotics facilitated the practice of 'best guess' medicine and of overprescription. But we no longer have the luxury of empirical therapy, and infectious-disease medicine must adopt twenty-first century innovations. Just as oncology has been transformed by fundamental research and the realization that cancer encompasses many diseases that require individualized therapies, so too must infection research. This requires a much greater understanding of the specific microorganisms that cause disease, their biology and their interactions with the host. As oncology has done already, the treatment of infection should adopt the model of personalized medicine, which would move the focus of the field from broad-spectrum agents to narrow, disease-specific antibiotics. The resulting narrow-spectrum era (Fig. 1) would have considerable impact on the next generation of antibiotics and generate fewer off-target effects on the

microbiome and decrease the pressure to evolve resistance. Such narrow-spectrum agents can be identified by mining the catalogue of discarded antibiotics — as evidenced by daptomycin — and through systematic testing of combinations of compounds that selectively target genus- and even species-specific pathways^{53,63}. The switch to narrow-spectrum agents should also expand the target base of antibiotics. Antibiotic discovery is considered to be target poor, with the drugs limited to targeting cell-wall synthesis, membrane integrity, translation, transcription and DNA synthesis. This might simply reflect common targets that are available to broad-spectrum agents. Narrow-spectrum agents offer a much larger vista of targets that are specific to bacteria and pathogenesis. For example, many of the processes that enable *in vivo* growth and virulence are targets for such antibiotics.

Because of time pressures and a lack of sensitive and reliable diagnostic agents, clinicians most often favour broad-spectrum agents. Advances in the accuracy of point-of-care diagnostics will be necessary to move forward the field of antibiotic discovery. And the business case for narrow-spectrum agents must address the consequences of lower sales volumes⁸¹. However, a spate of prizes for the development of point-of-care diagnostics, such as the UK Longitude Prize that seeks to tackle global antibiotic resistance, together with stunning advances in genome sequencing and informatics tools that can identify pathogens and their resistance profiles⁸² are creating the conditions necessary to support the discovery of narrow-spectrum antibiotics and their implementation through improved diagnostics. The narrow-spectrum era requires innovation, new economic models and collaboration across disciplines. Despite these challenges, the field of infectious-disease medicine can learn from other therapeutic areas to move from broad-spectrum empirical therapy that is no longer effective in the resistance era to the personalized and strategic approach of the emerging narrow-spectrum era. ■

Received 25 August; accepted 16 November 2015.

1. Fleming, A. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*. *Br. J. Exp. Pathol.* **10**, 226–236 (1929).
2. Comroe, J. H. Jr. Pay dirt: the story of streptomycin. Part I. From Waksman to Waksman. *Am. Rev. Respir. Dis.* **117**, 773–781 (1978).
3. Wright, G. D. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nature Rev. Microbiol.* **5**, 175–186 (2007).
4. Perry, J. A., Westman, E. L. & Wright, G. D. The antibiotic resistome: what's new? *Curr. Opin. Microbiol.* **21**, 45–50 (2014).
5. Forsberg, K. J. *et al.* Bacterial phylogeny structures soil resistomes across habitats. *Nature* **509**, 612–616 (2014).
6. Nesme, J. & Simonet, P. The soil resistome: a critical review on antibiotic resistance origins, ecology and dissemination potential in telluric bacteria. *Environ. Microbiol.* **17**, 913–930 (2015).
7. Finley, R. L. *et al.* The scourge of antibiotic resistance: the important role of the environment. *Clin. Infect. Dis.* **57**, 704–710 (2013).
8. Forsberg, K. J. *et al.* The shared antibiotic resistome of soil bacteria and human pathogens. *Science* **337**, 1107–1111 (2012).
9. D'Costa, V. M. *et al.* Antibiotic resistance is ancient. *Nature* **477**, 457–461 (2011).
This paper was the first to provide physical evidence of the ancient resistome, in 30,000-year-old permafrost.
10. Andersson, D. I. & Hughes, D. Persistence of antibiotic resistance in bacterial populations. *FEMS Microbiol. Rev.* **35**, 901–911 (2011).
11. Cox, G. & Wright, G. D. Intrinsic antibiotic resistance: mechanisms, origins, challenges and solutions. *Int. J. Med. Microbiol.* **303**, 287–292 (2013).
12. Fajardo, A. *et al.* The neglected intrinsic resistome of bacterial pathogens. *PLoS ONE* **3**, e1619 (2008).
An introduction to the concept of the intrinsic resistome, which is a target for antibiotic adjuvants.
13. Abraham, E. P. & Chain, E. An enzyme from bacteria able to destroy penicillin. *Nature* **146**, 837 (1940).
14. Davies, J. & Davies, D. Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* **74**, 417–433 (2010).
15. Fischbach, M. A. & Walsh, C. T. Antibiotics for emerging pathogens. *Science* **325**, 1089–1093 (2009).
16. Lewis, K. Antibiotics: recover the lost art of drug discovery. *Nature* **485**, 439–440 (2012).
17. Cho, H., Uehara, T. & Bernhardt, T. G. β -Lactam antibiotics induce a lethal malfunctioning of the bacterial cell wall synthesis machinery. *Cell* **159**, 1300–1311 (2014).
This paper provides evidence for the complexity of antibiotic-induced cell death.
18. Wilson, D. N. Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nature Rev. Microbiol.* **12**, 35–48 (2014).
19. Demirci, H. *et al.* A structural basis for streptomycin-induced misreading of the

- genetic code. *Nature Commun.* **4**, 1355 (2013).
20. Dwyer, D. J., Collins, J. J. & Walker, G. C. Unraveling the physiological complexities of antibiotic lethality. *Annu. Rev. Pharmacol. Toxicol.* **55**, 313–332 (2015).
 21. Macarron, R. *et al.* Impact of high-throughput screening in biomedical research. *Nature Rev. Drug Discov.* **10**, 188–195 (2011).
 22. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512 (1995).
 23. Kobayashi, K. *et al.* Essential *Bacillus subtilis* genes. *Proc. Natl Acad. Sci. USA* **100**, 4678–4683 (2003).
 24. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
 25. Forsyth, R. A. *et al.* A genome-wide strategy for the identification of essential genes in *Staphylococcus aureus*. *Mol. Microbiol.* **43**, 1387–1400 (2002).
 26. Jacobs, M. A. *et al.* Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA* **100**, 14339–14344 (2003).
 27. Jain, P. *et al.* Specialized transduction designed for precise high-throughput unmarked deletions in *Mycobacterium tuberculosis*. *mBio* **5**, e01245-14 (2014).
 28. Brown, E. D. & Wright, G. D. New targets and screening approaches in antimicrobial drug discovery. *Chem. Rev.* **105**, 759–774 (2005).
 29. Payne, D. J., Gwynn, M. N., Holmes, D. J. & Pompliano, D. L. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nature Rev. Drug Discov.* **6**, 29–40 (2007).
 30. Silver, L. L. Challenges of antibacterial discovery. *Clin. Microbiol. Rev.* **24**, 71–109 (2011).
 31. Tommasi, R., Brown, D. G., Walkup, G. K., Manchester, J. I. & Miller, A. A. ESKAPEing the labyrinth of antibacterial discovery. *Nature Rev. Drug Discov.* **14**, 529–542 (2015).
- An analysis of the challenges and failures of antibiotic discovery in 'big pharma'.**
32. Farha, M. A. & Brown, E. D. Unconventional screening approaches for antibiotic discovery. *Ann. NY Acad. Sci.* **1354**, 54–66 (2015).
 33. D'Elia, M. A., Pereira, M. P. & Brown, E. D. Are essential genes really essential? *Trends Microbiol.* **17**, 433–438 (2009).
 34. Joyce, A. R. *et al.* Experimental and computational assessment of conditionally essential genes in *Escherichia coli*. *J. Bacteriol.* **188**, 8259–8271 (2006).
 35. Nichols, R. J. *et al.* Phenotypic landscape of a bacterial cell. *Cell* **144**, 143–156 (2011).
 36. Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D. & Dalton, E. Simultaneous identification of bacterial virulence genes by negative selection. *Science* **269**, 400–403 (1995).
 37. Autret, N. & Charbit, A. Lessons from signature-tagged mutagenesis on the infectious mechanisms of pathogenic bacteria. *FEMS Microbiol. Rev.* **29**, 703–717 (2005).
 38. van Opijnen, T. & Camilli, A. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nature Rev. Microbiol.* **11**, 435–442 (2013).
 39. Turner, K. H., Wessel, A. K., Palmer, G. C., Murray, J. L. & Whiteley, M. Essential genome of *Pseudomonas aeruginosa* in cystic fibrosis sputum. *Proc. Natl Acad. Sci. USA* **112**, 4110–4115 (2015).
 40. Fahnoe, K. C. *et al.* Non-traditional antibacterial screening approaches for the identification of novel inhibitors of the glyoxylate shunt in Gram-negative pathogens. *PLoS ONE* **7**, e51732 (2012).
 41. Ziltni, S., Ferruccio, L. F. & Brown, E. D. Metabolic suppression identifies new antibacterial inhibitors under nutrient limitation. *Nature Chem. Biol.* **9**, 796–804 (2013).
- This paper describes a systematic screening approach for identifying antibacterial antimetabolites.**
42. Starkey, M. *et al.* Identification of anti-virulence compounds that disrupt quorum-sensing regulated acute and persistent pathogenicity. *PLoS Pathog.* **10**, e1004321 (2014).
 43. Pethe, K. *et al.* Discovery of Q203, a potent clinical candidate for the treatment of tuberculosis. *Nature Med.* **19**, 1157–1160 (2013).
 44. Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327**, 425–431 (2010).
 45. Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
 46. Butland, G. *et al.* Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **433**, 531–537 (2005).
 47. Babu, M. *et al.* Genetic interaction maps in *Escherichia coli* reveal functional crosstalk among cell envelope biogenesis pathways. *PLoS Genet.* **7**, e1002377 (2011).
 48. Gerdes, K., Christensen, S. K. & Løbner-Olesen, A. Prokaryotic toxin–antitoxin stress response loci. *Nature Rev. Microbiol.* **3**, 371–382 (2005).
 49. D'Elia, M. A. *et al.* Lesions in teichoic acid biosynthesis in *Staphylococcus aureus* lead to a lethal gain of function in the otherwise dispensable pathway. *J. Bacteriol.* **188**, 4183–4189 (2006).
 50. D'Elia, M. A. *et al.* Probing teichoic acid genetics with bioactive molecules reveals new interactions among diverse processes in bacterial cell wall biogenesis. *Chem. Biol.* **16**, 548–556 (2009).
 51. Swoboda, J. G. *et al.* Discovery of a small molecule that blocks wall teichoic acid biosynthesis in *Staphylococcus aureus*. *ACS Chem. Biol.* **4**, 875–883 (2009).
 52. Wang, H. *et al.* Discovery of wall teichoic acid inhibitors as potential anti-MRSA β -lactam combination agents. *Chem. Biol.* **20**, 272–284 (2013).
 53. Farha, M. A. *et al.* Inhibition of WTA synthesis blocks the cooperative action of PBPs and sensitizes MRSA to β -lactams. *ACS Chem. Biol.* **8**, 226–233 (2013).
 54. Farha, M. A. *et al.* Antagonism screen for inhibitors of bacterial cell wall biogenesis uncovers an inhibitor of undecaprenyl diphosphate synthase. *Proc. Natl Acad. Sci. USA* **112**, 11048–11053 (2015).
 55. Liu, A. *et al.* Antibiotic sensitivity profiles determined with an *Escherichia coli* gene knockout collection: generating an antibiotic bar code. *Antimicrob. Agents Chemother.* **54**, 1393–1403 (2010).
 56. Tan, C. M. *et al.* Restoring methicillin-resistant *Staphylococcus aureus* susceptibility to β -lactam antibiotics. *Sci. Transl. Med.* **4**, 126ra35 (2012).
 57. Pathania, R. *et al.* Chemical genomics in *Escherichia coli* identifies an inhibitor of bacterial lipoprotein targeting. *Nature Chem. Biol.* **5**, 849–856 (2009).
 58. Brynildsen, M. P., Winkler, J. A., Spina, C. S., Macdonald, I. C. & Collins, J. J. Potentiating antibacterial activity by predictably enhancing endogenous microbial ROS production. *Nature Biotechnol.* **31**, 160–165 (2013).
 59. Dwyer, D. J. *et al.* Antibiotics induce redox-related physiological alterations as part of their lethality. *Proc. Natl Acad. Sci. USA* **111**, E2100–E2109 (2014).
 60. Taylor, P. L., Rossi, L., De Pascale, G. & Wright, G. D. A forward chemical screen identifies antibiotic adjuvants in *Escherichia coli*. *ACS Chem. Biol.* **7**, 1547–1555 (2012).
 61. King, A. M. *et al.* Aspergillomarasmine A overcomes metallo- β -lactamase antibiotic resistance. *Nature* **510**, 503–506 (2014).
 62. Borisy, A. A. *et al.* Systematic discovery of multicomponent therapeutics. *Proc. Natl Acad. Sci. USA* **100**, 7977–7982 (2003).
 63. Ejim, L. *et al.* Combinations of antibiotics and nonantibiotic drugs enhance antimicrobial efficacy. *Nature Chem. Biol.* **7**, 348–350 (2011).
- The first demonstration of a systematic screen for antibiotic adjuvants.**
64. D'Costa, V. M., McGrann, K. M., Hughes, D. W. & Wright, G. D. Sampling the antibiotic resistome. *Science* **311**, 374–377 (2006).
 65. Bhullar, K. *et al.* Antibiotic resistance is prevalent in an isolated cave microbiome. *PLoS ONE* **7**, e34953 (2012).
 66. Bérdy, J. Bioactive microbial metabolites. *J. Antibiot.* **58**, 1–26 (2005).
 67. Baltz, R. H. Marcel Faber Roundtable: is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? *J. Ind. Microbiol. Biotechnol.* **33**, 507–513 (2006).
 68. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46**, 3–26 (2001).
 69. O'Shea, R. & Moser, H. E. Physicochemical properties of antibacterial compounds: implications for drug discovery. *J. Med. Chem.* **51**, 2871–2878 (2008).
 70. Kim, J., Kim, H. & Park, S. B. Privileged structures: efficient chemical 'navigators' toward unexplored biologically relevant chemical spaces. *J. Am. Chem. Soc.* **136**, 14629–14638 (2014).
 71. Galloway, W. R., Bender, A., Welch, M. & Spring, D. R. The discovery of antibacterial agents using diversity-oriented synthesis. *Chem. Commun.* **2009**, 2446–2462 (2009).
 72. Rachakonda, V., Alla, M., Kotipalli, S. S. & Ummanni, R. Design, diversity-oriented synthesis and structure activity relationship studies of quinolinyl heterocycles as antimycobacterial agents. *Eur. J. Med. Chem.* **70**, 536–547 (2013).
 73. Han, S., Zaniewski, R. P. & Marr, E. S. Structural basis for effectiveness of siderophore-conjugated monocarbams against clinically relevant strains of *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA* **107**, 22002–22007 (2010).
 74. Eisenstein, B. I., Oleson, F. B. & Baltz, R. H. Daptomycin: from the mountain to the clinic, with essential help from Francis Tally, MD. *Clin. Infect. Dis.* **50** (suppl. 1), S10–S15 (2010).
 75. Gerber, M. & Ackermann, G. OPT-80, a macrocyclic antimicrobial agent for the treatment of *Clostridium difficile* infections: a review. *Expert Opin. Investig. Drugs* **17**, 547–553 (2008).
 76. Lok, C. Mining the microbial dark matter. *Nature* **522**, 270–273 (2015).
 77. Wilson, M. C. *et al.* An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* **506**, 58–62 (2014).
 78. Ling, L. L. *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517**, 455–459 (2015).
 79. Brown, E. D. Is the GAIN Act a turning point in new antibiotic discovery? *Can. J. Microbiol.* **59**, 153–156 (2013).
 80. Spellberg, B., Bartlett, J., Wunderink, R. & Gilbert, D. N. Novel approaches are needed to develop tomorrow's antibacterial therapies. *Am. J. Respir. Crit. Care Med.* **191**, 135–140 (2015).
 81. Spellberg, B. The future of antibiotics. *Crit. Care* **18**, 228 (2014).
 82. McArthur, A. G. & Wright, G. D. Bioinformatics of antimicrobial resistance in the age of molecular epidemiology. *Curr. Opin. Microbiol.* **27**, 45–50 (2015).
 83. Theocharidis, A., van Dongen, S., Enright, A. J. & Freeman, T. C. Network visualization and analysis of gene expression data using BioLayout Express^{3D}. *Nature Protoc.* **4**, 1535–1550 (2009).

Acknowledgements The authors thank M. Farha for stimulating discussion with regards to the manuscript and S. French for creating Fig. 3. This work was supported by Discovery grants from the Natural Sciences and Engineering Research Council of Canada to E.D.B. (RGPIN 04384-2014) and G.D.W. (RGPIN 237480), by salary awards from the Canada Research Chairs Program to both E.D.B. and G.D.W., and by grants from the Canadian Institutes of Health Research to E.D.B. (MOP-81330 and MOP-15496) and G.D.W. (MT-13536).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing interests. Readers are welcome to comment on the online version of this paper at go.nature.com/srt9i. Correspondence should be addressed to E.D.B. (ebrown@mcmaster.ca) or G.D.W. (wrightge@mcmaster.ca).

The functional diversity of retinal ganglion cells in the mouse

Tom Baden^{1,2,3*}, Philipp Berens^{1,2,3,4,5*}, Katrin Franke^{1,2,3,6*}, Miroslav Román Rosón^{1,2,3,6}, Matthias Bethge^{1,2,5,7} & Thomas Euler^{1,2,3}

In the vertebrate visual system, all output of the retina is carried by retinal ganglion cells. Each type encodes distinct visual features in parallel for transmission to the brain. How many such ‘output channels’ exist and what each encodes are areas of intense debate. In the mouse, anatomical estimates range from 15 to 20 channels, and only a handful are functionally understood. By combining two-photon calcium imaging to obtain dense retinal recordings and unsupervised clustering of the resulting sample of more than 11,000 cells, here we show that the mouse retina harbours substantially more than 30 functional output channels. These include all known and several new ganglion cell types, as verified by genetic and anatomical criteria. Therefore, information channels from the mouse eye to the mouse brain are considerably more diverse than shown thus far by anatomical studies, suggesting an encoding strategy resembling that used in state-of-the-art artificial vision systems.

Visual processing begins in the retina (reviewed in ref. 1). Here, photoreceptors feed into bipolar cells², which provide input to a diverse set of retinal ganglion cells (RGCs). Each type of RGC tiles the retinal surface and extracts specific features of the visual scene for transmission to the brain. However, it is still unclear how many such parallel retinal ‘feature channels’ exist, and what they encode.

Early studies classified cells into ON, OFF or ON–OFF and transient or sustained types (for example, refs 3, 4) based on the response of individual RGCs to light stimulation. These studies also identified RGC types selective for local motion, motion direction or uniform illumination^{3,5–7}. In the most complete physiological survey to date, Farrow and Masland⁸ clustered ~450 mouse RGCs by their light responses into at least 12 functional types using multi-electrode array (MEA) recordings, suggesting a similar number of feature channels in the retina. In contrast, anatomical classifications of RGC dendritic morphologies estimated around 15–20 types (for example, refs 9–12). Recently, Sümbül and co-workers¹⁰ found at least 16 types using unsupervised clustering together with genetic markers. If each of these anatomically distinct types performed one function, there should be no more than ~20 retinal output channels.

Commonly, RGCs of the same ‘genuine’ type are thought to share the same physiology, morphology, intra-retinal connectivity, retinal mosaic, immunohistochemical profile and genetic markers. Whether these features suffice to define a type and how classification schemes should be organized is a matter of long-standing debate^{13–16}. For example, if also axonal projections were considered type-specific, this could result in a much greater variety of retinal output channels. In zebrafish, RGCs show at least 50 unique combinations of ‘dendro-axonal RGC morphologies’ targeting a total of ten anatomically defined projection fields¹⁷. RGCs in mice project to at least 40 targets¹⁸, suggesting that there may be an even larger number of mouse RGC types.

Reliably recording from all RGC types

Here, we sought to test this idea and determine the number of functional output channels of the mouse retina, to obtain a complete picture

of what the mouse’s eye tells the mouse’s brain. We used two-photon Ca^{2+} imaging to record light-evoked activity in all cells within a patch of the ganglion cell layer (GCL). Cells were loaded with the fluorescent Ca^{2+} indicator Oregon-Green BAPTA-1 (OGB-1) by bulk electroporation¹⁹ (Fig. 1a). This approach resulted in near-complete (>92%) staining of GCL cells, with less than 1% damaged cells²⁰. To acquire a patch of several hundreds of cells, we recorded up to nine neighbouring $110 \times 110 \mu\text{m}$ fields (at 7.8 Hz), each containing 80 ± 20 GCL somata (Fig. 1a and Supplementary Video 1). In total, more than 11,000 cells were sampled.

We presented four light stimuli (Fig. 1b): (i) a full-field ‘chirp’ stimulus to characterize polarity, kinetics and the preference for temporal frequencies and contrasts; (ii) a moving bar to probe for direction and orientation selectivity; (iii) binary dense noise to estimate receptive fields; and (iv) an alternating long- or short-wavelength (‘green/blue’) full-field stimulus to probe for chromatic preference (Methods). This set of stimuli was chosen to cover a large stimulus space that distinguishes different response types.

We performed recordings in the ventral retina, as verified by the mean chromatic response preference^{21,22} of each imaged field (Extended Data Fig. 1e), to control for retinotopic sources of variability^{23–25}. In addition, we always presented stimuli at the same light levels in the low photopic regime (Methods).

Our approach allowed us to determine the soma size and position of each recorded cell. Immunohistochemistry (Methods) was used in a subset of experiments: displaced amacrine cells (dACs), which are GABAergic and/or cholinergic and represent a substantial fraction of GCL somata²⁶, were labelled for GAD67 and ChAT (Fig. 1c; $n = 522$ of 1,584 cells; double-labelled: $n = 96$). Melanopsin-labelling^{27,28} identified strongly melanopsin-expressing intrinsically photosensitive RGCs (ipRGCs, $n = 18$ of 905 cells). SMI-32 labelled a small set of GCL cells ($n = 85$ of 1,912 cells), including starburst dACs²⁹, and was used to identify alpha RGCs (large, strongly labelled somata^{9,23}). In addition, we performed recordings in two transgenic lines, Pvalb (PV) and Pcp2 (Methods; Fig. 1d), to relate individual recorded cells to known genetically defined populations.

¹Bernstein Centre for Computational Neuroscience, 72076 Tübingen, Germany. ²Centre for Integrative Neuroscience, University of Tübingen, 72076 Tübingen, Germany. ³Institute for Ophthalmic Research, 72076 Tübingen, Germany. ⁴Baylor College of Medicine, Houston, Texas 77030, USA. ⁵Institute of Theoretical Physics, University of Tübingen, 72076 Tübingen, Germany. ⁶Graduate Training Centre of Neuroscience, University of Tübingen, 72074 Tübingen, Germany. ⁷Max Planck Institute of Biological Cybernetics, 72076 Tübingen, Germany.

*These authors contributed equally to this work.

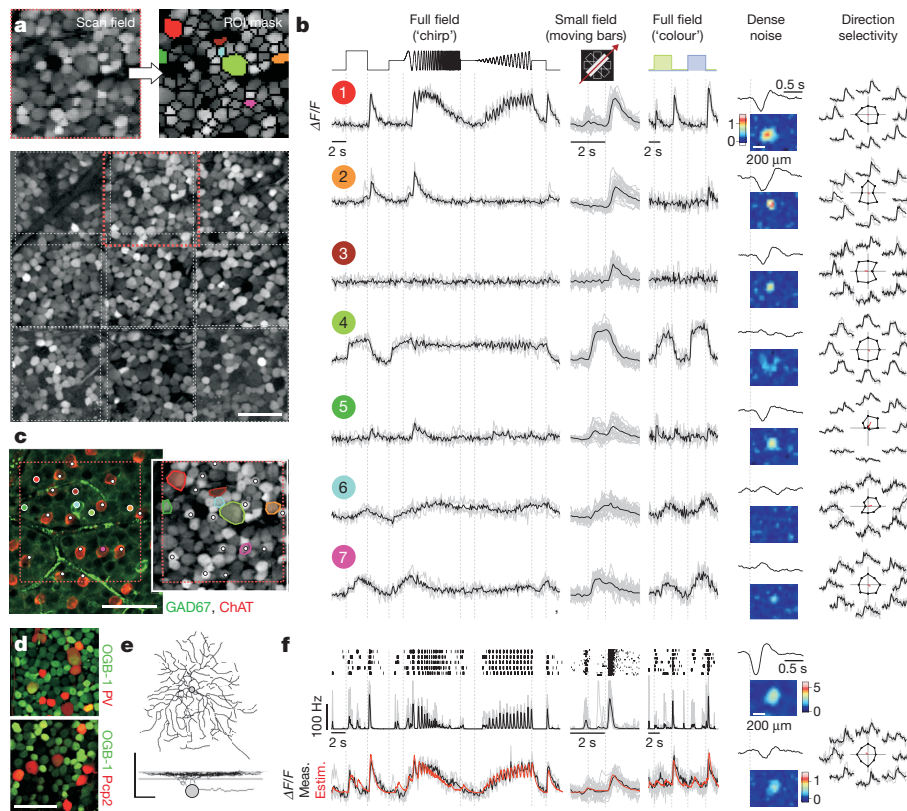


Figure 1 | Data collection. **a**, Whole-mounted mouse retina, electroporated with OGB-1 and recorded with a two-photon microscope (64×64 pixels at 7.8 Hz) in the GCL. Scan fields (left; $110 \times 110 \mu\text{m}$) comprised 80 ± 20 cells. Regions of interest (ROIs) (right), were placed semi-automatically. Bottom, montage of nine consecutively recorded fields (rectangles; top panels indicated by red dashed line). **b**, Ca^{2+} signals from seven regions of interest colour-coded in **a**. Single trials in grey, averages of $n = 4$ (chirp, green/blue) or 24 (moving bars) trials in black. Responses to four visual stimuli: full-field chirp, bright bars moving in eight directions, full-field alternating green/blue and binary noise for space-time kernels. Rightmost column, direction- and orientation-selectivity: traces by motion direction; polar plot of peak response, vector sum in red.

Finally, we made electrical single-cell recordings from RGCs ($n = 84$) followed by dye filling to reconstruct their dendritic morphology (Fig. 1e). For all cells with Ca^{2+} and spike activity recorded simultaneously ($n = 17$), Ca^{2+} responses estimated from spike trains closely resembled measured Ca^{2+} responses (Fig. 1f; Extended Data Fig. 1a–d).

A probabilistic clustering framework

Combining locally complete optical population recordings, genetic and immunohistochemical labels, as well as electrical measurements, yielded a comprehensive data set of GCL light responses to a set of standardized visual stimuli. This provided the unprecedented opportunity for an unbiased characterization of the retinal output. Since the data set (11,210 cells, $n = 50$ retinas; Extended Data Fig. 2) is too complex to be interpreted manually (for discussion, ref. 30), we used a clustering approach, making our analysis as objective and quantitative as possible.

In the first step, we used an automatic unsupervised clustering procedure to identify response prototypes of GCL cells (Extended Data Fig. 2d, f; Methods). Specifically, we used sparse principal component analysis (sPCA) to extract features from the light-driven Ca^{2+} signals of the GCL cells (Methods), which identified many classically used temporal response features such as ON and OFF responses with different kinetics or selectivity to different temporal frequencies. We then used a Mixture of Gaussian model on this feature set for clustering (Methods). In the second step, we post-processed the clustered data

c, Left, experiment in **(a)** immunostained for GAD67 (green; GABAergic ACs) and ChAT (red; starburst ACs). Right, from **(a)**; both images show same colour-coded regions of interest (left, dots; right, region of interest outlines) and starburst ACs (white dots): cell 6 is GAD67-positive, cell 7 is a starburst AC. **d**, OGB-1 (green) electroporated retina from transgenic mice with tdTomato (red) expressed in sets of RGCs (top, PV (Pvalb); bottom, Pcp2). **e**, **f**, Simultaneous Ca^{2+} imaging and electrical recording: dye-filled, anatomically reconstructed cell (**e**, top, whole-mount; bottom, profile, lines mark ChAT bands). Light responses (**f**) from top to bottom: spike raster and rate (20-ms bins), recorded (black) and reconstructed (orange) Ca^{2+} signal. Scale bars: $50 \mu\text{m}$ unless otherwise indicated.

to make it accessible for interpretation, including the identification of clusters corresponding to dAC types based on GAD67 staining and isolation of alpha RGCs⁹ with large somata from similarly responding cells (Extended Data Fig. 2g, h–j). The validity of this step was verified in detail below (see section ‘Example RGC types’).

Finally, we arranged the clusters according to a hierarchical tree based on their functional similarity (Methods) and suggest a grouping scheme based on available domain knowledge (Fig. 2a–c). Some groups span different branches, as the tree was solely based on the functional response features.

This framework yielded a total of 46 groups ($n = 7,982$ cells, 71.2% of all cells, Extended Data Fig. 2h), divided into 32 RGC groups ($n = 5,024$, 62.9% of grouped cells; including 4 groups, G_{29-32} , from ‘uncertain’ clusters) and 14 dAC groups ($n = 2,958$, 37.1% of grouped cells). The estimated fraction of dACs (between 37.1% and 50.6%, if including all uncertain groups) is within the expected range²⁶. We did not analyse dACs in detail (see Extended Data Fig. 3a–c; Supplementary Video 2, Supplementary Figures 1: 50–75, and Supplementary Discussion).

A minimum of 32 mouse RGC types

The identified 32 RGC groups comprised non-direction-selective (9 OFF, 12 ON, 3 ON–OFF) and direction-selective (2 ON–OFF, 4 ON, 2 OFF) groups (Fig. 2a–c; Supplementary Figures 2: 1–32) and accounted for all known RGC types in the mouse retina¹⁵. This includes groups corresponding to three alpha types³¹ ($G_{5,8,24}$,

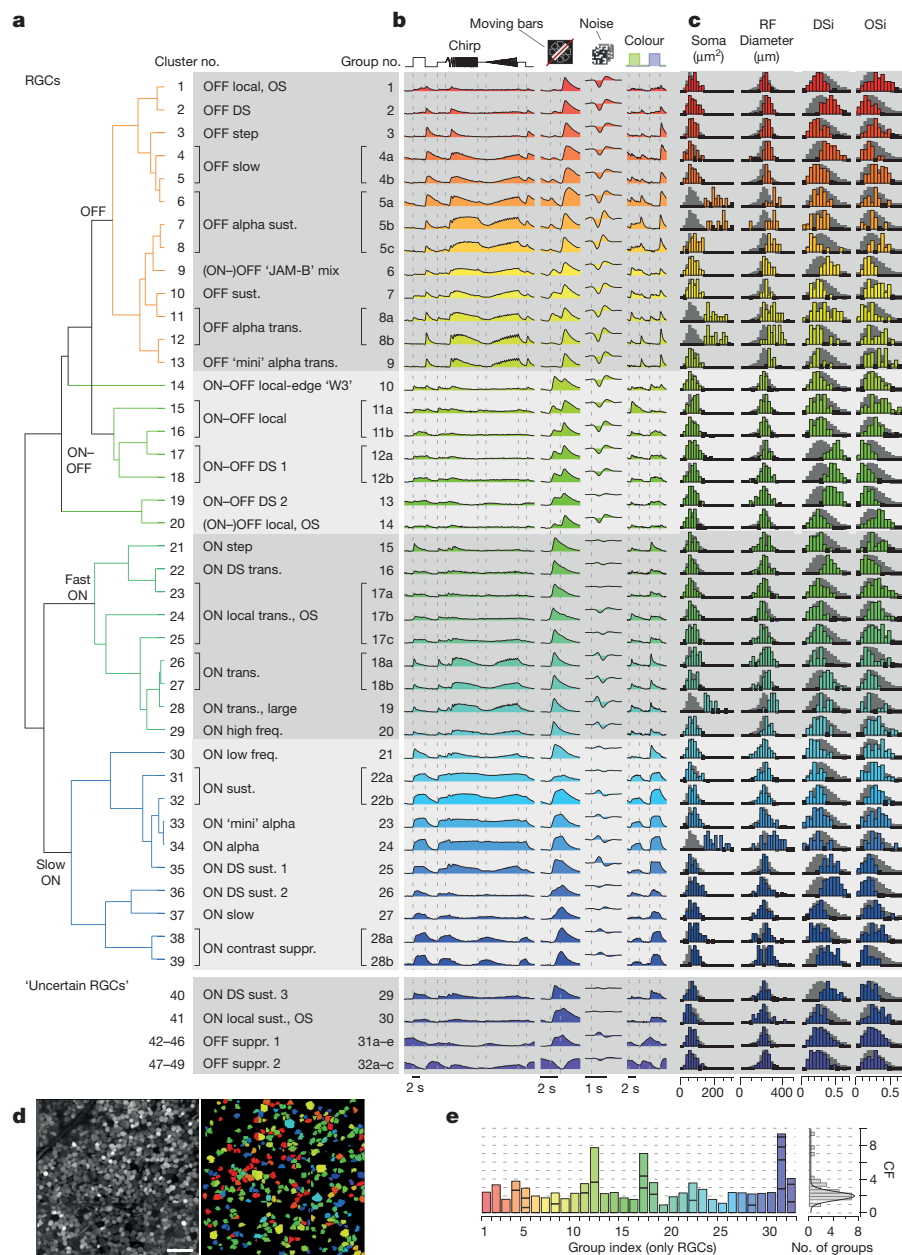


Figure 2 | Functional RGC types of the mouse retina. **a**, Cluster-dendrogram (Methods) with groups indicated: $n = 28$ RGC and $n = 4$ 'uncertain' RGC groups. **b**, Cluster-mean Ca^{2+} responses to the four stimuli. **c**, Selected metrics, from left to right: region of interest (soma) area, receptive field (RF) diameter (2 s.d. of Gaussian), direction-selectivity index (DSi) and orientation-selectivity index (Osi) (Methods). Background-histograms demarcate all RGCs. **d**, Experiment (left, from Fig. 1a, bottom) with RGCs colour-coded by group (right). dACs and discarded cells not shown. **e**, Coverage factor (CF) calculated from receptive field area of RGC groups, with horizontal divisions delineating individual clusters (left) and distribution of coverage factors across groups (right). Scale bar in **d**, 50 μm .

see also below), ON-OFF³² ($G_{12,13}$) and ON direction-selective types³³ ($G_{16,25,26,29}$), JAM-B (ref. 24) (G_6) and W3 cells²⁵ (G_{10}), and the OFF 'suppressed-by-contrast' cell³⁴ ($G_{31,32}$). The allocation of cells to individual groups was uniform across space (Fig. 2d) and the fraction of the population accounted for by broad response types was consistent across experiments (Extended Data Fig. 3e).

As each RGC type is thought to tile the retina, we calculated each group's functional coverage factor based on its average receptive field size and its relative abundance (Extended Data Fig. 4; Methods). A single RGC type would yield a coverage factor of ~ 1 without receptive field overlap, and a coverage factor of ~ 2 with 30% overlap. A coverage factor $\ll 1$ may indicate that a type has been artificially split.

The average coverage factor across all RGC groups was 2.0 ± 0.7 (mean \pm s.d. of Gaussian fit, Fig. 2e), broadly consistent with reported coverage factors for mouse RGCs (roughly 2–3; see Supplementary Table 1 and Supplementary Discussion). Coverage factors higher than two may indicate groups consisting of multiple types. For example, G_{12} corresponding to ON-OFF direction-selective cells³², has a coverage factor of 7.7, consistent with four ON-OFF direction-selective types, each preferring a different motion direction³⁵ (see below).

The mixed non-direction-selective groups G_{17} and G_{31} probably contain more than one type, as supported by multiple distinct morphologies and genetic identities (for example, $G_{31,32}$, Extended Data Fig. 5) or response properties (for example, G_{17} , see below).

Taken together, our coverage factor analysis suggests that the number of unique functional RGC types in the mouse is substantially above 32, probably as high as 40, in particular since classical ipRGCs (that is, M1) as well as at least one PV-positive small-field RGC type were largely discarded based on their low signal-to-noise ratio (S/N) response to our stimuli (see Supplementary Discussion). This is about three times the highest number of physiologically defined RGC types to date⁸ and about twice the highest anatomical diversity reported in mouse¹¹ (Supplementary Table 2).

Example RGC types

There are three types of alpha cells known in the mouse retina³¹: the sustained (G_5) and transient OFF alpha (G_8), and the ON alpha (G_{24}). These cells are characterized by their large somata, a feature that we used during the post-processing step (Extended Data Fig. 2i–j). For the transient OFF and the ON alpha we found similarly responding cells

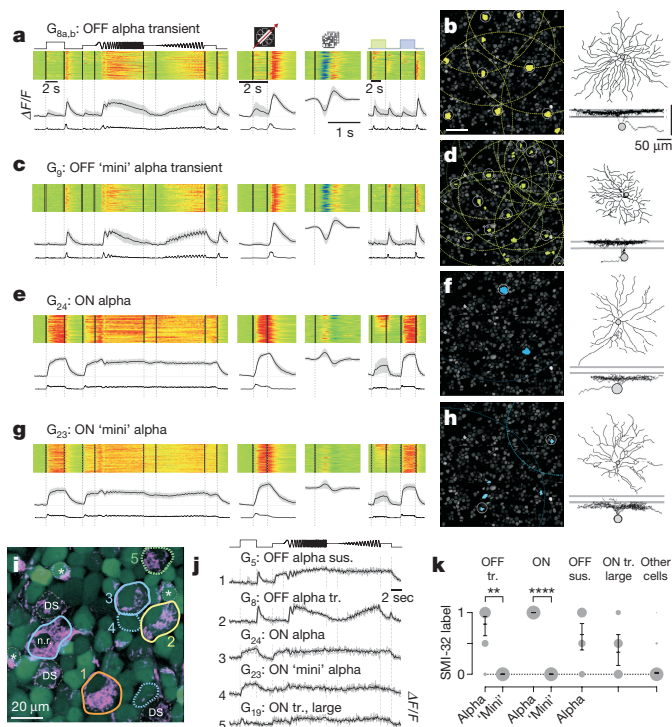


Figure 3 | Classical alpha RGCs and their 'mini' counterparts.

a, b, Functional 'fingerprint' of classical transient OFF alpha cells ($G_{8a,b}$). **a**, Light-evoked Ca^{2+} responses of $n = 80$ cells: heat maps (top) of individual responses, response averages (with 1 s.d.) and firing rates estimated from Ca^{2+} signals below (cf. Extended Data Fig. 1). **b**, Left, $G_{8a,b}$ somata (yellow) and receptive fields (dotted) indicated in example experiment. Grey circles mark cells with receptive fields above quality criterion (Methods). Right, sample morphology of a $G_{8a,b}$ cell filled after electrical single-cell recording. For details, see Supplementary Figure 2: 8. **c**, G_9 RGCs ($n = 68$), dubbed OFF mini alpha transient because of their similarity in light response to $G_{8a,b}$ RGCs (compare with **a**). **e, f**, G_{24} RGCs ($n = 44$), identified as classical ON alpha. **g, h**, G_{23} RGCs ($n = 113$), dubbed ON mini alpha (compare with **e**). Mini alphas have smaller receptive field diameters than classical alphas (median in μm , 95% confidence interval): G_9 , 280 (270–293) versus G_8 , 306 (294–315) with $P = 0.01232$, and G_{23} , 236 (218–256) versus G_{24} , 319 (290–352) with $P = 0.00026$ (rank-sum test). **i**, overlay of OGB-1-stained cells (green) and SMI-32 (magenta). SMI-32-positive RGCs include classical alphas (solid contours; n.r. indicates a non-responsive cell), one large-soma non-alpha cell (green, dotted contour) as well as weakly-labelled ON–OFF direction-selective cells (dashed contours) and starburst ACs (asterisks). Mini alpha cells (blue, dotted contours) are SMI-32-negative. **j**, Chirp-evoked Ca^{2+} responses for five cells in **i**. **k**, SMI-32 statistics (OFF tr.: alpha, $n = 16$, mini, $n = 3$; ON: alpha, $n = 6$, mini, $n = 15$; OFF sus. alpha, $n = 14$; ON tr. large, $n = 7$; other, $n = 957$; means with 95% confidence intervals; ** $P \leq 0.01$; **** $P \leq 0.0001$, logistic regression).

with smaller somata, which we named transient OFF 'mini' (G_9) and ON 'mini' alpha cells (G_{23}), respectively (Fig. 3a–h). For the sustained OFF alpha (G_5) we did not identify an obvious 'mini' version.

We tested whether these pairs consisted of distinct types using SMI-32 immunohistochemistry on a subset of cells ($n = 1,912$, Fig. 3i–k). These immunolabels were not used during post-processing (Extended Data Fig. 2g). In an example field, all 'classical' alpha cells were SMI-32-positive (numbers 1–3, Fig. 3i, j). In contrast, two ON mini alphas (no. 4) had smaller somata and were SMI-32-negative, despite a response profile similar to the ON alpha. Across our entire set of stained cells, transient OFF and ON alpha cells were consistently SMI-32-positive, while their respective mini counterparts were not (Fig. 3k, $P < 0.01$, logistic regression, Methods). In addition, ON alphas, unlike their mini counterparts, were PV-positive ($n_{\text{ONalpha}} = 16/26$ versus $n_{\text{ONalpha-mini}} = 1/37$, Supplementary Figures 2: 23, 24). Finally, alpha and mini

alpha types formed independent mosaics (Fig. 3b, d, f, h). Transient OFF mini alpha cells ($n = 2$) stratified similarly to transient OFF alpha cells ($n = 5$), but with smaller dendritic arbors, consistent with their smaller receptive fields (Fig. 2c; Supplementary Figures 2: 8, 9; for statistics, see Fig. 3 legend). The same was the case for ON alpha and mini alpha RGCs (Supplementary Figures 2: 23, 24). Together, these data suggest that alphas and mini alphas are distinct cell types. The OFF sustained alpha was only moderately SMI-32-immunoreactive (Fig. 3k), consistent with previous reports²³. An additional, weakly SMI-32-positive RGC group with large somata (no. 5 in Fig. 3i, j) was dubbed 'ON transient, large' (G_{19}).

The classical 'local edge detector' (ref. 36) or W3 cell²⁵ probably corresponds to G_{10} (Extended Data Fig. 6a–f). This type had small somata and receptive fields, and responded poorly to full-field stimuli. Instead, G_{10} cells reliably responded to moving bars with a brisk burst at the leading and trailing edges, consistent with previous reports²⁵. Intracellular filling confirmed the cell's morphological identity (Extended Data Fig. 6c). A second group, G_{11} , had similar response properties with a reduced leading-edge response, potentially corresponding to a W3 variant²⁵.

Contrast-suppressed RGCs (for example, see refs 37, 38) occur in many species and a type of OFF suppressed-by-contrast (Sbc) cell has recently been described in mice³⁴ (G_{31}). We found a new ON Sbc cell type (G_{28}), which responded slowly to a full-field increase in light level but was suppressed by temporal contrast (Extended Data Fig. 6g–i). One candidate morphology diffusively stratified across the entire inner plexiform layer (IPL), while another stratified exclusively in the ON layer.

Direction and orientation selectivity

Our RGC groups contained $n = 1,757$ direction-selective cells (35% of RGCs; Extended Data Fig. 7a–e; Methods). Most direction-selective cells (70%) were sorted into eight groups ($G_{2,6,12,13,16,25,26,29}$). This high functional diversity among direction-selective cells is in agreement with studies of direction-selective RGC-specific transgenic mouse lines³⁹. Further direction-selective clusters were grouped with functionally similar non-direction-selective counterparts ($G_{4,18,28}$), as they contained less cells than the matching non-direction-selective clusters. These may include RGCs with slightly asymmetric dendritic arbors (for example, ref. 10) that could lead to directional bias in response to motion. Classically, these cells are not considered direction-selective.

To identify the motion axes of direction-selective groups, we registered the orientation of the retina in a subset of experiments ($n = 3,830$ cells, $n = 12$ retinas, Fig. 4; Extended Data Fig. 7f, g). Three direction-selective groups largely preferred one direction: the ON direction-selective transient (G_{16}), the JAM-B RGC²⁴ (G_6) and one of the sustained ON direction-selective cells (G_{25}) preferred backward, upward and forward motion, respectively. The other direction-selective groups responded to more than one direction. The two groups of ON–OFF direction-selective cells ($G_{12,13}$) contain the four classical subtypes³², with room for additional ones, as suggested by genetic analyses and central projections (for example, ref. 40; Supplementary Discussion). The three classical sustained ON direction-selective types³³ are included in $G_{25,26,29}$.

We found two OFF direction-selective groups ($G_{2,6}$), with G_6 containing JAM-B RGCs²⁴. These cells responded weakly to ON and strongly to OFF light changes, consistent with the JAM-B's polarity switch with stimulus size. In addition, we identified a new OFF direction-selective cell (G_2 , Extended Data Fig. 6m–r). This group did not co-stratify with the ChAT bands, which play a role in most known direction-selective circuits (reviewed in ref. 39). Morphology, response profile, and directional preference of G_2 suggests that these cells are not the JAM-B RGCs (see above).

As predicted by recordings from mouse lateral geniculate nucleus⁴¹ and superior colliculus⁴², we identified $n = 730$ (14.5%) orientation-selective RGCs (Fig. 4b). While most RGC groups contained only few

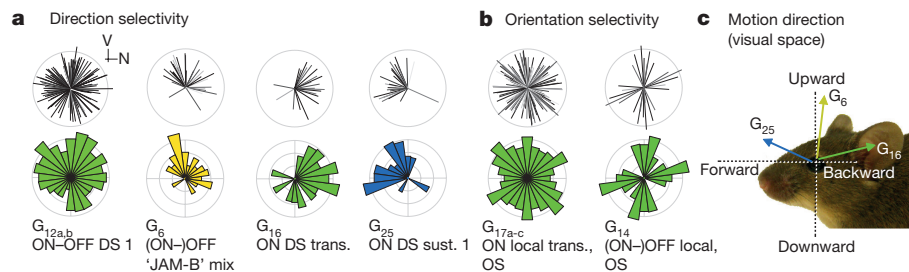


Figure 4 | Direction and orientation selectivity. **a**, Pairs of retinocentric polar plots showing distributions of preferred motion directions of selected direction-selective (DS) RGC groups (V, ventral; N, nasal). Top plot of each pair: preferred directions, with length representing direction-selective index and grey level p_{DS} (Methods). Bottom plot of

each pair: circular area-normalized histogram. **b**, As for **a**, but for selected orientation-selective (OS) RGCs. Further direction-selective/orientation-selective groups detailed in Extended Data Fig. 7. **c**, Motion directions in the visual space of the mouse.

orientation-selective cells, $G_{1,14,17,30}$ contained disproportionately more ($\sim 30\%$ each). (ON-)OFF orientation-selective cells (G_{14}) were selective for vertical and horizontal orientations, whereas ON transient orientation-selective cells (G_{17}) included many preferred orientations, consistent with its coverage factor of seven. Additional experiments with contrast inversed moving and stationary bars ($n = 826$ cells) revealed further functional diversity in response to these stimuli among G_{17} cells (Extended Data Fig. 7h–s; Supplementary Discussion).

Genetic and anatomical RGC types

To link RGC groups to genetically-defined populations, we performed a subset of our experiments in the PV⁴³ and Pcp2 (ref. 44) transgenic mouse lines (Extended Data Figs 5, 8; $n_{PV} = 173$ cells in 24 retinas, $n_{Pcp2} = 15$ cells in 3 retinas). PV- and Pcp2-positive cells were sorted into 20 (PV_{a-t}) and 6 (Pcp2_{a-t}) groups (Extended Data Fig. 5b–d), with 14 and 3 of the groups containing $n \geq 3$ genetically labelled cells, respectively. In the case of the PV line, many matches were very robust (for example, the ON alpha cells: G_{24} or PV_r or 'PV1' from ref. 45). Nevertheless, the count of ≥ 14 functional types in the PV line is much higher than the previously described eight PV types⁴⁵, suggesting higher functional diversity than appreciated in earlier studies (Supplementary Discussion).

Next, we used our comparatively small sample of $n = 84$ morphologically reconstructed cells to link functional groups to anatomically defined types. In many cases, it was possible to identify likely morphologies for functional RGC groups (Extended Data Fig. 9; Supplementary Discussion). To generate an approximate mapping of functional groups to dendritic depth profiles in the IPL (Fig. 5), we averaged the stratification profiles of all reconstructed cells, weighted by the correlation coefficient between each cell's light response and the functional group average (to full-field chirp and moving bars; Methods).

The resulting map reproduced many known principles of inner retinal organization. For example, OFF (G_{1-9}) and sustained ON ($G_{21-24,27,28,30}$) groups mostly stratified in the upper and lower half of the IPL⁴ and groups with more transient responses closer to the centre

(for example, $G_{1,2,8-10,20}$)^{46,47}. Most groups corresponding to known types had expected stratification profiles (for example, W3, ON-OFF direction-selective and classical alphas) with few exceptions (for example, $G_{6,11,18,19}$). Our prediction was for the JAM-B (G_6) to stratify mainly below the ChAT band, while the cell is known to stratify above it, as it also does in our individual examples (Extended Data Fig. 9a). This is probably caused by many cell types with responses comparable to those of the JAM-B confounding the G_6 IPL profile.

Conclusions

We found that a minimum of 32 different functional types of RGCs could be distinguished based on their light responses and basic anatomical criteria. The unusually high abundance of some of these functional types and evidence from immunohistochemistry suggests that further sub-divisions are needed. Accordingly, the number of distinct visual feature channels available to the mouse brain appears to be two- to threefold that of previous estimates.

Taken together, our RGC groups cover a broad range of 'classical' features such as polarity, receptive field size, frequency and contrast sensitivity (Extended Data Fig. 10). In particular, RGC groups broadly span feature dimensions such as response polarity and their preference for global versus local stimuli. Less balanced is the temporal frequency selectivity, with only a few groups preferring high frequencies, particularly for groups with low contrast preference.

We verified our suggested functional classification by showing that: (i) most functionally defined types (ii) exhibited a similar coverage factor; (iii) some could be linked to genetically defined populations; and (iv) types had consistent morphology/dendritic stratification profiles in the IPL. Nonetheless, our definitions certainly remain incomplete; the classification of mouse RGCs will need to be refined by the expansion of the probed stimulus space, the use of cell-type-selective genetic lines⁴⁰ or single-cell transcriptomics⁴⁸, and integration with data from large-scale electron microscopy⁴⁹. However, even our comparatively basic analysis already reveals a large diversity in feature coding by mouse RGCs, very different from how digital cameras encode images, rather

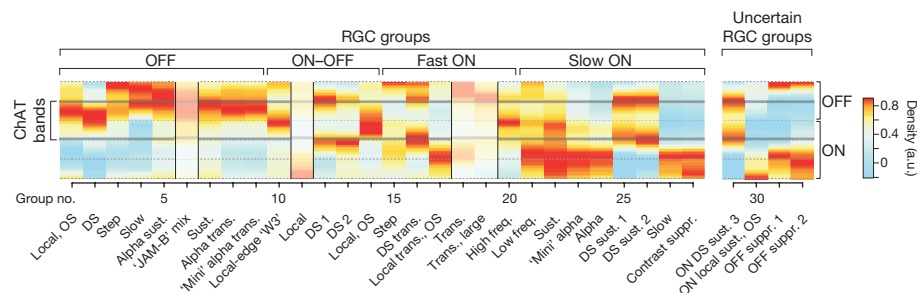


Figure 5 | Mapping RGC groups to morphologies. Heat map of each RGC group's estimated dendritic stratification across the IPL (compare with Figure 2); ON/OFF sublaminae and ChAT bands indicated. Warmer colours represent higher dendritic densities (Methods). Shaded IPL

profiles indicate deviation from known stratification pattern (G_6) or an unexpected pattern given a potentially novel group's response polarity ($G_{11,18,19}$). a.u., arbitrary units, DS, direction-selective; OS, orientation-selective.

resembling an encoding strategy used in state-of-the-art artificial vision systems⁵⁰. In the future, the ‘fingerprint’ of different functional RGC types introduced here can provide a frame of reference for systematic investigations of feature coding by RGCs and for detecting functional changes in degenerated retina.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 February; accepted 17 November 2015.

Published online 6 January 2016.

- Masland, R. H. The neuronal organization of the retina. *Neuron* **76**, 266–280 (2012).
- Euler, T., Haverkamp, S., Schubert, T. & Baden, T. Retinal bipolar cells: elementary building blocks of vision. *Nature Rev. Neurosci.* **15**, 507–519 (2014).
- Lettvin, J., Maturana, H., McCulloch, W. & Pitts, W. What the frog’s eye tells the frog’s brain. *Proc. IRE* **47**, 1940–1951 (1959).
- Werblin, F. S. & Dowling, J. E. Organization of the retina of the mudpuppy, *Necturus maculosus*. II. Intracellular recording. *J. Neurophysiol.* **32**, 339–355 (1969).
- Cleland, B. G. & Levick, W. R. Brisk and sluggish concentrically organized ganglion cells in the cat’s retina. *J. Physiol. (Lond.)* **240**, 421–456 (1974).
- Barlow, H. B., Hill, R. M. & Levick, W. R. Rabbit retinal ganglion cells responding selectively to direction and speed of image motion in the rabbit. *J. Physiol. (Lond.)* **173**, 377–407 (1964).
- Devries, S. H. & Baylor, D. A. Mosaic arrangement of ganglion cell receptive fields in rabbit retina. *J. Neurophysiol.* **78**, 2048–2060 (1997).
- Farrow, K. & Masland, R. H. Physiological clustering of visual channels in the mouse retina. *J. Neurophysiol.* **105**, 1516–1530 (2011).
- Coombs, J., van der List, D., Wang, G.-Y. & Chalupa, L. M. Morphological properties of mouse retinal ganglion cells. *Neuroscience* **140**, 123–136 (2006).
- Sümbül, U. *et al.* A genetic and computational approach to structurally classify neuronal types. *Nature Commun.* **5**, 3512 (2014).
- Völgyi, B., Chheda, S. & Bloomfield, S. A. Tracer coupling patterns of the ganglion cell subtypes in the mouse retina. *J. Comp. Neurol.* **512**, 664–687 (2009).
- Kong, J.-H., Fish, D. R., Rockhill, R. L. & Masland, R. H. Diversity of ganglion cells in the mouse retina: unsupervised morphological classification and its limits. *J. Comp. Neurol.* **489**, 293–310 (2005).
- Rowe, M. H. & Stone, J. Naming of neurones. Classification and naming of cat retinal ganglion cells. *Brain Behav. Evol.* **14**, 185–216 (1977).
- Seung, H. S. & Sümbül, U. Neuronal cell types and connectivity: lessons from the retina. *Neuron* **83**, 1262–1272 (2014).
- Sanes, J. R. & Masland, R. H. The types of retinal ganglion cells: current status and implications for neuronal classification. *Annu. Rev. Neurosci.* **38**, 221–246 (2015).
- Rodieck, R. W. & Brening, R. K. Retinal ganglion cells: properties, types, genera, pathways and trans-species comparisons. *Brain Behav. Evol.* **23**, 121–164 (1983).
- Robles, E., Laurell, E. & Baier, H. The retinal projectome reveals brain-area-specific visual representations generated by ganglion cell diversity. *Curr. Biol.* **24**, 2085–2096 (2014).
- Morin, L. P. & Studholme, K. M. Retinofugal projections in the mouse. *J. Comp. Neurol.* **522**, 3733–3753 (2014).
- Briggman, K. L. & Euler, T. Bulk electroporation and population calcium imaging in the adult mammalian retina. *J. Neurophysiol.* **105**, 2601–2609 (2011).
- Euler, T. *et al.* Eyecup scope-optical recordings of light stimulus-evoked fluorescence signals in the retina. *Pflügers Arch.* **457**, 1393–1414 (2009).
- Wang, Y. V., Weick, M. & Demb, J. B. Spectral and temporal sensitivity of cone-mediated responses in mouse retinal ganglion cells. *J. Neurosci.* **31**, 7670–7681 (2011).
- Baden, T. *et al.* A tale of two retinal domains: near-optimal sampling of achromatic contrasts in natural scenes through asymmetric photoreceptor distribution. *Neuron* **80**, 1206–1217 (2013).
- Bleckert, A., Schwartz, G. W., Turner, M. H., Rieke, F. & Wong, R. O. L. Visual space is represented by nonmatching topographies of distinct mouse retinal ganglion cell types. *Curr. Biol.* **24**, 310–315 (2014).
- Kim, I. J., Zhang, Y., Yamagata, M., Meister, M. & Sanes, J. R. Molecular identification of a retinal cell type that responds to upward motion. *Nature* **452**, 478–482 (2008).
- Zhang, Y., Kim, I. J., Sanes, J. R. & Meister, M. The most numerous ganglion cell type of the mouse retina is a selective feature detector. *Proc. Natl Acad. Sci. USA* **109**, E2391–E2398 (2012).
- Schlamp, C. L. *et al.* Evaluation of the percentage of ganglion cells in the ganglion cell layer of the rodent retina. *Mol. Vis.* **19**, 1387–1396 (2013).
- Berson, D. M., Castrucci, A. M. & Provencio, I. Morphology and mosaics of melanopsin-expressing retinal ganglion cell types in mice. *J. Comp. Neurol.* **518**, 2405–2422 (2010).
- Ecker, J. L. *et al.* Melanopsin-expressing retinal ganglion-cell photoreceptors: Cellular diversity and role in pattern vision. *Neuron* **67**, 49–60 (2010).
- Lim, E.-J., Kim, I.-B., Oh, S.-J. & Chun, M.-H. Identification and characterization of SMI32-immunoreactive amacrine cells in the mouse retina. *Neurosci. Lett.* **424**, 199–202 (2007).
- Armañanzas, R. & Ascoli, G. A. Towards the automatic classification of neurons. *Trends Neurosci.* **38**, 307–318 (2015).
- van Wyk, M., Wässle, H. & Taylor, W. R. Receptive field properties of ON- and OFF-ganglion cells in the mouse retina. *Vis. Neurosci.* **26**, 297–308 (2009).
- Weng, S., Sun, W. & He, S. Identification of ON-OFF direction-selective ganglion cells in the mouse retina. *J. Physiol. (Lond.)* **562**, 915–923 (2005).
- Sun, W., Deng, Q., Levick, W. R. & He, S. ON direction-selective ganglion cells in the mouse retina. *J. Physiol. (Lond.)* **576**, 197–202 (2006).
- Tien, N.-W., Pearson, J. T., Heller, C. R., Demas, J. & Kerschensteiner, D. Genetically identified suppressed-by-contrast retinal ganglion cells reliably signal self-generated visual stimuli. *J. Neurosci.* **35**, 10815–10820 (2015).
- Oyster, C. W. & Barlow, H. B. Direction-selective units in rabbit retina: distribution of preferred directions. *Science* **155**, 841–842 (1967).
- Levick, W. R. Receptive fields and trigger features of ganglion cells in the visual streak of the rabbits retina. *J. Physiol. (Lond.)* **188**, 285–307 (1967).
- Sivyer, B., Taylor, W. R. & Vaney, D. I. Uniformity detector retinal ganglion cells fire complex spikes and receive only light-evoked inhibition. *Proc. Natl Acad. Sci. USA* **107**, 5628–5633 (2010).
- Nikolaev, A., Leung, K. M., Odermatt, B. & Lagnado, L. Synaptic mechanisms of adaptation and sensitization in the retina. *Nature Neurosci.* **16**, 934–941 (2013).
- Vaney, D. I., Sivyer, B. & Taylor, W. R. Direction selectivity in the retina: symmetry and asymmetry in structure and function. *Nature Rev. Neurosci.* **13**, 194–208 (2012).
- Rivlin-Etzion, M. *et al.* Transgenic mice reveal unexpected diversity of on-off direction-selective retinal ganglion cell subtypes and brain structures involved in motion processing. *J. Neurosci.* **31**, 8760–8769 (2011).
- Zhao, X., Chen, H., Liu, X. & Cang, J. Orientation-selective responses in the mouse lateral geniculate nucleus. *J. Neurosci.* **33**, 12751–12763 (2013).
- Feinberg, E. H. & Meister, M. Orientation columns in the mouse superior colliculus. *Nature* **519**, 229–232 (2015).
- Hippenmeyer, S. *et al.* A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS Biol.* **3**, e159 (2005).
- Lewis, P. M., Gritti-Linde, A., Smeyne, R., Kottmann, A. & McMahon, A. P. Sonic hedgehog signaling is required for expansion of granule neuron precursors and patterning of the mouse cerebellum. *Dev. Biol.* **270**, 393–410 (2004).
- Farrow, K. *et al.* Ambient illumination toggles a neuronal circuit switch in the retina and visual perception at cone threshold. *Neuron* **78**, 325–338 (2013).
- Roska, B. & Werblin, F. Vertical interactions across ten parallel, stacked representations in the mammalian retina. *Nature* **410**, 583–587 (2001).
- Baden, T., Berens, P., Bethge, M. & Euler, T. Spikes in mammalian bipolar cells support temporal layering of the inner retina. *Curr. Biol.* **23**, 48–52 (2013).
- Macosko, E. Z. *et al.* highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
- Denk, W. & Horstmann, H. Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol.* **2**, e329 (2004).
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25* (eds Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q.) 1097–1105 (Curran Associates, Inc., 2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank G. Eske for technical support and H. S. Seung, P. R. Martin, and A. S. Tolias for discussion. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) (Werner Reichardt Centre for Integrative Neuroscience Tübingen, EXC 307 to M.B. and T.E.; BA 5283/1-1 to T.B.; BE 5601/1-1 to P.B.), the German Federal Ministry of Education and Research (BMBF) (BCCN Tübingen, FKZ 01GQ1002 to M.B. and T.E.), the BW-Stiftung (AZ 1.16101.09 to T.B.), the intramural fortune program of the University of Tübingen (2125-0-0 to T.B.) and the National Institute of Neurological Disorders and Stroke of the National Institutes of Health (U01NS090562 to T.E.).

Author Contributions T.B., P.B., M.B. and T.E. designed the study; K.F. performed imaging experiments with help from T.B.; K.F. and M.R.R. performed electrophysiological experiments with help from T.B.; T.B., P.B., K.F. and M.R.R. performed pre-processing; P.B. developed the clustering framework with the help of M.B.; T.B. and P.B. analysed the data with input from T.E.; T.B., P.B. and T.E. wrote the manuscript.

Author Information Data (original data, clustering and grouping results) as well as Matlab code for visualization are available from <http://www.retinal-functomics.org>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.E. (thomas.euler@cin.uni-tuebingen.de).

METHODS

No statistical methods were used to predetermine sample size.

Animals and tissue preparation. All procedures were performed in accordance with the law on animal protection issued by the German Federal Government (Tierschutzgesetz) and approved by the institutional animal welfare committee of the University of Tübingen. For all experiments, we used 4- to 8-week-old mice of either sex. In addition to C57Bl6 (wild-type) mice, we used the transgenic lines Pvalb^{Cre} (PV; JAX 008069, The Jackson Laboratory; ref. 43), Pcp2^{Cre} (Pcp2; JAX 006207; ref. 44) and ChAT^{Cre} (JAX 006410; ref. 51), cross-bred with the red fluorescence Cre-dependent reporter line Ai9:tdTomato (JAX 007905). Owing to the exploratory nature of our study, we did not use randomization and blinding.

Animals were housed under a standard 12 h day/night rhythm. For activity recordings, animals were dark-adapted for ≥ 1 h, then anaesthetized with isoflurane (Baxter) and killed by cervical dislocation. The eyes were enucleated and hemisected in carboxygenated (95% O₂, 5% CO₂) artificial cerebral spinal fluid (ACSF) solution containing (in mM): 125 NaCl, 2.5 KCl, 2 CaCl₂, 1 MgCl₂, 1.25 NaH₂PO₄, 26 NaHCO₃, 20 glucose, and 0.5 L-glutamine (pH 7.4). Bulk electroporation of the fluorescent Ca²⁺ indicator Oregon-Green BAPTA-1 (OGB-1) into the ganglion cell layer (GCL) was carried out as described before^{19,47}. In brief, the retina was dissected from the eyecup, flat-mounted onto an Anodisc (#13, 0.2 μ m pore size, GE Healthcare) with the GCL facing up, and placed between two 4-mm horizontal plate electrodes (CUY700P4E/L, Nepagene/Xceltis). A 10 μ l drop of 5 mM OGB-1 (hexapotassium salt; Life Technologies) in ACSF was suspended from the upper electrode and lowered onto the retina. After application of 10–12 pulses (+9 V, 100 ms pulse width, at 1 Hz) from a pulse generator/wide-band amplifier combination (TGP110 and WA301, Thurlby handar/Farnell), the tissue was moved to the recording chamber of the microscope, where it was continuously perfused with carboxygenated ACSF at $\sim 37^\circ\text{C}$ and left to recover for ~ 60 min before the recordings started. In all experiments with wild-type mice, ACSF contained $\sim 0.1 \mu\text{M}$ Sulforhodamine-101 (SR101, Invitrogen) to reveal blood vessels and any damaged cells in the red fluorescence channel²⁰. All procedures were carried out under very dim red (>650 nm) light.

Two-photon Ca²⁺ imaging and light stimulation. We used a MOM-type two-photon microscope (designed by W. Denk, MPI, Martinsried; purchased from Sutter Instruments/Science Products). Design and procedures were described previously²⁰. In brief, the system was equipped with a mode-locked Ti:Sapphire laser (MaiTai-HP DeepSee, Newport Spectra-Physics) tuned to 927 nm, two fluorescence detection channels for OGB-1 (HQ 510/84, AHF/Chroma) and SR101 (HQ 630/60, AHF), and a water immersion objective (W Plan-Apochromat 20x/1.0 DIC M27, Zeiss). For image acquisition, we used custom-made software (ScanM, by M. Müller, MPI, Martinsried, and T.E.) running under IGOR Pro 6.3 for Windows (Wavemetrics), taking 64 \times 64 pixel image sequences (7.8 frames per s) for activity scans or 512 \times 512 pixel images for high-resolution morphology scans.

For light stimulation, we focused a DLP projector (K11, Acer) through the objective, fitted with band-pass-filtered light-emitting diodes (LEDs) ('green', 578 BP 10; and 'blue', HC 405 BP 10, AHF/Chroma) that roughly match the spectral sensitivity of mouse M- and S-opsins. LEDs were synchronized with the microscope's scan retracing. Stimulator intensity (as photoisomerisation rate, 10³ P* per s per cone) was calibrated as described previously⁵² to range from 0.6 and 0.7 (black image) to 18.8 and 20.3 for M- and S-opsins, respectively. Owing to two-photon excitation of photopigments, an additional, steady illumination component of $\sim 10^4$ P* per s per cone was present during the recordings (for detailed discussion, see ref. 22). For all experiments, the tissue was kept at a constant intensity level (see stimuli below) for at least 30 s after the laser scanning started before light stimuli were presented. Four types of light stimulus were used (Fig. 1b, top): (i) full-field 'chirp' stimulus consisting of a bright step and two sinusoidal intensity modulations, one with increasing frequency and one with increasing contrast; (ii) 0.3 \times 1 mm bright bar moving at 1 mm s⁻¹ in eight directions¹⁹; (iii) alternating blue and green 3-s flashes; and (iv) binary dense noise, a 20 \times 15 matrix with 40 μ m pixel-side length; each pixel displayed an independent, perfectly balanced random sequence at 5 Hz yielding a total running time of 5 min for receptive field (RF) mapping. In some experiments, we used in addition dark moving bars (like (ii) but contrast-inversed), and stationary bright or dark 0.2 \times 0.8 mm bars flashed for 1 s in six orientations (see Extended Data Fig. 7h–s). Except for (iii), stimuli were achromatic, with matched photo-isomerization rates for M- and S-opsins.

Single-cell electrophysiology, dye filling and morphological reconstruction. TdTomato- or OGB-1-labelled RGCs were targeted using two-photon imaging for juxtacellular recordings using borosilicate electrodes (4–6 M Ω) filled with ACSF with added SR101 (250 μM). Data were acquired using an Axoclamp-900A or Axopatch 200A amplifier in combination with a Digidata 1440 (all: Molecular Devices), digitized at 10 kHz and analysed offline using IGOR Pro. We presented the same light stimuli as for the Ca²⁺ imaging. In some experiments, electrical recordings and Ca²⁺ imaging was performed simultaneously. After the recording,

the membrane under the electrode was opened using a voltage 'buzz' to let the cell fill with dye by diffusion for approximately 30 min; then two-photon image stacks were acquired to document the cell's morphology.

Filled cells were traced semi-automatically offline using the Simple Neurite Tracer plugin implemented in Fiji (<http://fiji.sc/Fiji>), yielding cell skeletons. If necessary, we used the original image stack to correct the skeletons for any warping of the IPL using custom-written scripts in IGOR Pro. To this end, we employed the SR101-stained blood vessel plexuses on either side of the IPL as landmarks to define the IPL borders (see below). Only cells where the filling quality allowed full anatomical reconstruction were used for analysis (see below).

Immunohistochemistry. Following Ca²⁺ imaging, retinas were mounted onto filter paper (0.8 μ m pore size, Millipore) and fixed in 4% paraformaldehyde (in PBS) for 15 min at 4°C. Immunolabelling was performed using antibodies against ChAT (choline-acetyltransferase; goat anti-ChAT, 1:100, AB144P, Millipore), GAD67 (glutamate decarboxylase; mouse anti-GAD67, 1:100, MAB5046, Millipore), SMI-32 (mouse anti-SMI32, 1:100, SMI-32R-100, Covance), and melanopsin (rabbit anti-melanopsin, 1:4000, AB-N38, Advanced Targeting Systems) for 4 days. Secondary antibodies were Alexa Fluor conjugates (1:750, 16 h, Life Technologies). For each retina, the recorded region was identified by the local blood vessel pattern and confirmed by comparing size and position of individual somata in the GCL. Image stacks were acquired on a confocal microscope (Nikon Eclipse C1) equipped with a $\times 60$ oil objective (1.4 NA). The degree of immunolabelling of GCL cells was evaluate and rated (from 0, negative, to 4 positive) using z-stacks. Attribution of labelled somata to recorded ones was performed manually using ImageJ (<http://imagej.nih.gov/ij>) and IGOR Pro.

Data analysis. Data analysis was performed using Matlab 2012 and 2014a (The Mathworks Inc.), and IGOR Pro. Data were organized in a custom written schema using the DataJoint for Matlab framework (<http://datajoint.github.io/>; D. Yatsenko, Tolias lab, Baylor College of Medicine).

Pre-processing. Regions of interest (ROIs), corresponding to somata in the GCL, were defined semi-automatically by custom software (D. Velychko, CIN) based on a high-resolution (512 \times 512 pixels) image stack of the recorded field. Then, the Ca²⁺ traces for each ROI were extracted (as $\Delta F/F$) using the IGOR Pro-based image analysis toolbox SARFIA (<http://www.igorexchange.com/project/SARFIA>). A stimulus time marker embedded in the recording data served to align the Ca²⁺ traces relative to the visual stimulus with a temporal precision of 2 ms. Stimulus-aligned Ca²⁺ traces for each ROI were imported into Matlab for further analysis.

First, we de-trended the Ca²⁺ traces by high-pass filtering above ~ 0.1 Hz. For all stimuli except the dense noise (for RF mapping), we then subtracted the baseline (median of first eight samples), computed the median activity $r(t)$ across stimulus repetitions (typically three to five repetitions) and normalized it such that $\max_t(|r(t)|) = 1$.

Receptive field mapping. We mapped the linear RFs of the neurons by computing the Ca²⁺ transient-triggered average. To this end, we resampled the temporal derivative of the Ca²⁺ response $\dot{c}(t)$ at 10-times the stimulus frequency and used Matlab's findpeaks function to detect the times t_i at which Ca²⁺ transients occurred. We set the minimum peak height to 1 s.d., where the s.d. was robustly estimated using:

$$\sigma = \frac{\text{median}(|\dot{r}(t)|)}{0.6745}$$

We then computed the Ca²⁺ transient-triggered average stimulus, weighting each sample by the steepness of the transient:

$$F(x, y, \tau) = \frac{1}{M} \sum_{i=1}^M \dot{c}(t_i) S(x, y, t_i + \tau)$$

Here, $S(x, y, t)$ is the stimulus, τ is the time lag (ranging from approximately -320 to $1,380$ ms) and M is the number of Ca²⁺ events. We smoothed this raw RF estimate using a 5×5 pixel Gaussian window for each time lag separately. RF maps shown correspond to a s.d. map, where the s.d. is calculated over time lags τ :

$$F_{\text{map}}(x, y) = \text{s.d.}[F(x, y, \tau)]_{\tau}$$

To extract the RF's position and scale, we fitted it with a 2D Gaussian function using Matlab's lsqcurvefit. The time course of the receptive field $F_{\text{ic}}(\tau)$ was estimated by the average of the eight pixels closest to the fitted RF centre (according to the Mahalanobis distance) weighted by a Gaussian profile. RF quality (Q_{RF}) was measured as one minus the fraction of variance explained by the Gaussian fit \bar{F}_{map} ,

$$Q_{\text{RF}} = 1 - \frac{\text{Var}[F_{\text{map}} - \bar{F}_{\text{map}}]}{\text{Var}[F_{\text{map}}]}$$

Direction and orientation selectivity. To extract time course and directional tuning of the Ca^{2+} response to the moving bar stimulus, we performed a singular value decomposition (SVD) on the T by D normalized mean response matrix M (times samples by number of directions; $T = 32$; $D = 8$; Extended Data Fig. 7a, b):

$$[U, S, V] = \text{svd}(M)$$

This procedure decomposes the response into a temporal component in the first column of U and a direction dependent component or tuning curve in the first column of V , such that the response matrix can be approximated as an outer product of the two:

$$M \approx S_{11} U_{:,1} V_{:,1}^T$$

An advantage of this procedure is that it does not require manual selection of time bins for computing direction tuning, but extracts the direction tuning curve given the varying temporal dynamics of different neurons.

To measure direction selectivity (DS) and its significance, we projected the tuning curve $V_{:,1}$ on a complex exponential $\phi_k = \exp(i\alpha_k)$, where α_k is the direction in the k th condition:

$$K = \phi^T V_{:,1}$$

This is mathematically equivalent to computing the vector sum in the 2D plane or computing the power in the first Fourier component. We computed a DS index as the resulting vector length

$$\text{DSi} = |K|$$

correcting for the direction spacing. We additionally assessed the statistical significance of direction tuning using a permutation test⁵³. To this end, we created surrogate trials (that is, stimulus repetitions) by shuffling the trial labels (that is, destroying any relationship between condition and response), computed the tuning curve for each surrogate trial and projected it on the complex exponential ϕ . Carrying out the procedure 1,000 times generated a null distribution for K , assuming no direction tuning. We used the percentile of the true K as the P value for direction tuning (Extended Data Fig. 7c). Importantly, a large DSi does not necessarily result in a small P value, for example, in the case of large trial to trial variability. As a result, the DSi distributions of significantly and not significantly direction tuned neurons show substantial overlap (Extended Data Fig. 7d, e). Therefore, a simple threshold as a DS criterion (for example, $\text{DSi} > 0.4$) does not provide a good separation into direction selective cell types and others.

Orientation selectivity (OS) was assessed in an analogous way. However, we used the complex exponential $\phi_k = \exp(2i\alpha_k)$, corresponding to the second Fourier component.

Response quality index. To measure how well a cell responded to a stimulus (chirp, moving bar, colour), we computed the signal-to-noise ratio

$$QI = \frac{\text{Var}[\langle C \rangle_r]_t}{\langle \text{Var}[C]_t \rangle_r}$$

where C is the T by R response matrix (time samples by stimulus repetitions) and $\langle \rangle_x$ and $\text{Var}[\]_x$ denote the mean and variance across the indicated dimension, respectively. If all trials are identical such that the mean response is a perfect representative of the response, QI is equal to 1. If all trials are completely random with fixed variance (so that the mean response is not informative about the individual trial responses at all), QI is proportional to $1/R$.

For further analysis, we used only cells that responded well to the chirp and/or the moving bar stimulus ($QI_{\text{chirp}} > 0.45$ or $QI_{\text{DS}} > 0.6$).

Full-field index. The full-field index was computed as

$$\text{FFi} = \frac{QI_{\text{DS}} - QI_{\text{chirp}}}{QI_{\text{DS}} + QI_{\text{chirp}}}$$

comparing the response quality to a local stimulus (moving bar) and a global stimulus (chirp).

ON-OFF index. ON-OFF preference was measured as

$$\text{OOi} = \frac{\langle r_{\text{ON}} \rangle_t - \langle r_{\text{OFF}} \rangle_t}{\langle r_{\text{ON}} \rangle_t + \langle r_{\text{OFF}} \rangle_t}$$

where r_{ON} and r_{OFF} are defined as the activity during the response to the leading edge of the moving bar (the first 400 ms of the ON response) and the trailing edge of the moving bar (the first 400 ms of the OFF response).

Colour selectivity index. Colour selectivity was measured for the ON response using

$$\text{GB}_{\text{ON}} = \frac{\max(r_{\text{ON,green}}^2) - \max(r_{\text{ON,blue}}^2)}{\max(r_{\text{ON,green}}^2) + \max(r_{\text{ON,blue}}^2)}$$

and for the OFF response using an analogous definition. Here, $r_{\text{ON,green}}$ and $r_{\text{ON,blue}}$ are the responses in a time window of 1,280 ms after onset of the green and blue stimulus, respectively.

Feature extraction. We used sparse principle component analysis⁵⁴, as implemented in the SpaSM toolbox by K. Sjöström *et al.* (<http://www2.imm.dtu.dk/projects/spasm/>), to extract sparse response features from the responses to the chirp, colour, and moving bar stimulus, resulting in features which use only a small number of time bins. The extracted features are localized in time and therefore readily interpretable (for example, 'high-frequency feature'), although this constraint was not explicitly enforced by the algorithm (Extended Data Fig. 2e). We also explored alternative feature extraction techniques such as regular PCA, but these resulted in inferior cluster quality. In addition, they required manually defining regions corresponding to specific parts of the stimulus (for example, frequency chirp) to yield localized and interpretable features.

We extracted 20 features with 10 non-zero time bins from the mean response to the chirp (averaging across trials) and 6 features with 10 non-zero time bins from the mean response to the colour stimulus. For the moving bar stimulus, we extracted 8 features with 5 non-zero time bins from the response time course (see above) and 4 features with 6 non-zero time bins from its temporal derivative. All features were in the temporal domain, ensuring spatial invariance. In addition, we used two features from the time course of the RF, extracted with regular PCA. Overall, this procedure resulted in a 40 dimensional feature vector for each cell. Before clustering, we standardized each feature separately across the population of cells.

Clustering. DS and non-DS cells were clustered independently, classifying cells as DS if the permutation test resulted in $P < 0.05$ (see above). We fit each data set with a Mixture of Gaussians model using the expectation-maximization algorithm (Matlab's `gmdistribution` object). We constrained the covariance matrix for each component to be diagonal, resulting in 80 parameters per component (40 for the mean, 40 for the variances). We further regularized the covariance matrix by adding a constant (10^{-5}) to the diagonal. To find the optimal number of clusters, we evaluated the Bayesian information criterion⁵⁵

$$\text{BIC} = -2 \log[L] + M \log[N]$$

where L is the log-likelihood of the model, N is the number of cells and M is the number of parameters in the model, that is, $M = 81C - 1$ where C is the number of clusters and the contributions that arose from means, variances and mixture proportions (which have to add to 1). Although other choices such as the Akaike information criterion (AIC) would have been possible, we found the BIC to yield a good compromise between model complexity and quality, since the AIC is known to find too many clusters for large sample sizes. We also computed log Bayes factors as $2\Delta\text{BIC}$ for each candidate cluster number to test how strong the evidence for further splitting is. Values > 6 were treated as strong evidence in favour of further splitting. The minimum of the BIC coincided well with the number of clusters after which there was no strong evidence for further adding more clusters. To avoid local minima, we restarted the EM algorithm 20 times per candidate number of clusters and used the solution with the largest likelihood. This procedure resulted in 24 and 48 clusters for DS and non-DS cells, respectively (Extended Data Fig. 2a).

To evaluate cluster quality, we rank-ordered the posterior probabilities for cluster assignment for each cluster, normalized for cluster size and averaged across clusters for non-DS and DS cells separately (Extended Data Fig. 2b). The steep decays of the sigmoidal functions indicate good cluster separability. To check how consistent the clustering was against subsampling of the data, we created 20 surrogate data sets containing random selections of 90% of the cells. We fit these surrogate data sets with a Mixture of Gaussians model with the optimal number of clusters determined on the original data set. For each cluster mean in these models, we computed the correlations with the most similar cluster for the model fit on the original data set. To summarize the similarity of clusterings, we computed the median correlation across clusters (Extended Data Fig. 2c). On average, the clusterings obtained on the surrogate and the original data set were very similar (mean median correlation: 0.96 ± 0.19 and 0.97 ± 0.01 ; mean \pm s.d.; for DS and non-DS cells separately).

In addition, we performed an alternative clustering version, where we did not split the data in DS and non-DS cells but added DSi, OSi, soma and receptive field size as features. The identified clusters were very similar, but this strategy failed to identify most DS types as separate clusters, except for the ON-OFF DS

cell. Therefore, we decided to first isolate significant DS cells and cluster them separately, before merging similarly responding DS and non-DS clusters (see below), if we did not find a reason to keep the DS group as a separate RGC type. Nevertheless, a strategy equally justified as ours could start with the alternative clustering and then split those clusters containing large fractions of DS cells.

Automatic identification of RGCs and ACs. A subset of cells was stained against GAD67 to identify dACs (see above). The intensity of this staining was manually rated as follows: -2 (absent), -1 (probably absent), 0 (uncertain), 1 (probably present), and 2 (present). For each cluster, we computed the average staining from the labelled cells (average number of cells with GAD67 information per cluster: 16.8 ± 10.0 , mean \pm s.d.). Clusters with an average staining < -0.2 were labelled RGCs ($n = 30$ clusters), those with average staining > 0.2 were labelled ACs ($n = 26$). Clusters with average staining in-between those values ($n = 5$), or those that contained 6 or less cells with GAD67 information ($n = 8$) were labelled as uncertain, unless other clear criteria such as soma size or genetic labels indicated that they are ACs or RGCs. In this case they were manually allocated to RGC or AC ($n = 3$ and $n = 2$, respectively). Two clusters automatically classified as AC were included in the uncertain group due to their functional similarity with the OFF-suppressed types (G_{31}). This procedure resulted in 33 RGC clusters, 10 uncertain clusters and 26 AC clusters.

Identification of alpha-RGCs. We extracted all cells with large cell bodies ($> 136 \mu\text{m}^2$; mean \pm 1 s.d. of total soma size distribution; Extended Data Fig. 2i, j) from RGC and uncertain clusters. Predominantly, these cells had been assigned to nine of the clusters. We re-clustered those cells using a Mixture of Gaussians model as described above, resulting in 16 clusters (Extended Data Fig. 2j). Receptive field size was not used in this process. Five of these clusters could be clearly associated with the three known alpha-RGC types and their response profiles³¹ (trans. OFF alpha, 2; sust. OFF alpha, 2; ON alpha, 1). Cells in these clusters were SMI-32-positive, as expected from alpha RGCs (Fig. 3i, k). Probably, this procedure missed some alpha cells, as somata close to the edge of scan fields were cut and we thus underestimate the soma size of these cells (for example, G_{5c} , see Fig. 2a–c). Remaining cells were kept in their original cluster. Logistic regression was used to assess the effect cell type (alpha vs. mini) on SMI-32 staining (absent vs. present). We used the Matlab implementation `fitglm` with a binomial nonlinearity. 95%-confidence intervals on the proportion of SMI32-positive cells were computed using bootstrapping with 1,000 samples.

Dendrogram. We used a standard linkage algorithm on the means of the RGC groups in the standardized feature space with correlation distance $d_{ij} = 1 - \text{corr}(m_i, m_j)$ and average unweighted distance and plotted the result as a dendrogram (using Matlab functions `linkage` and `dendrogram`). The leaf order was optimized using the Matlab function `optimalleaforder` and modified for clarity of presentation.

Calculation of coverage factor. We calculated each group's coverage factor (CF)

$$\text{CF} = \frac{n_{\text{Cells}} \cdot A_{\text{RF,median}}}{A_{\text{Scan}}}$$

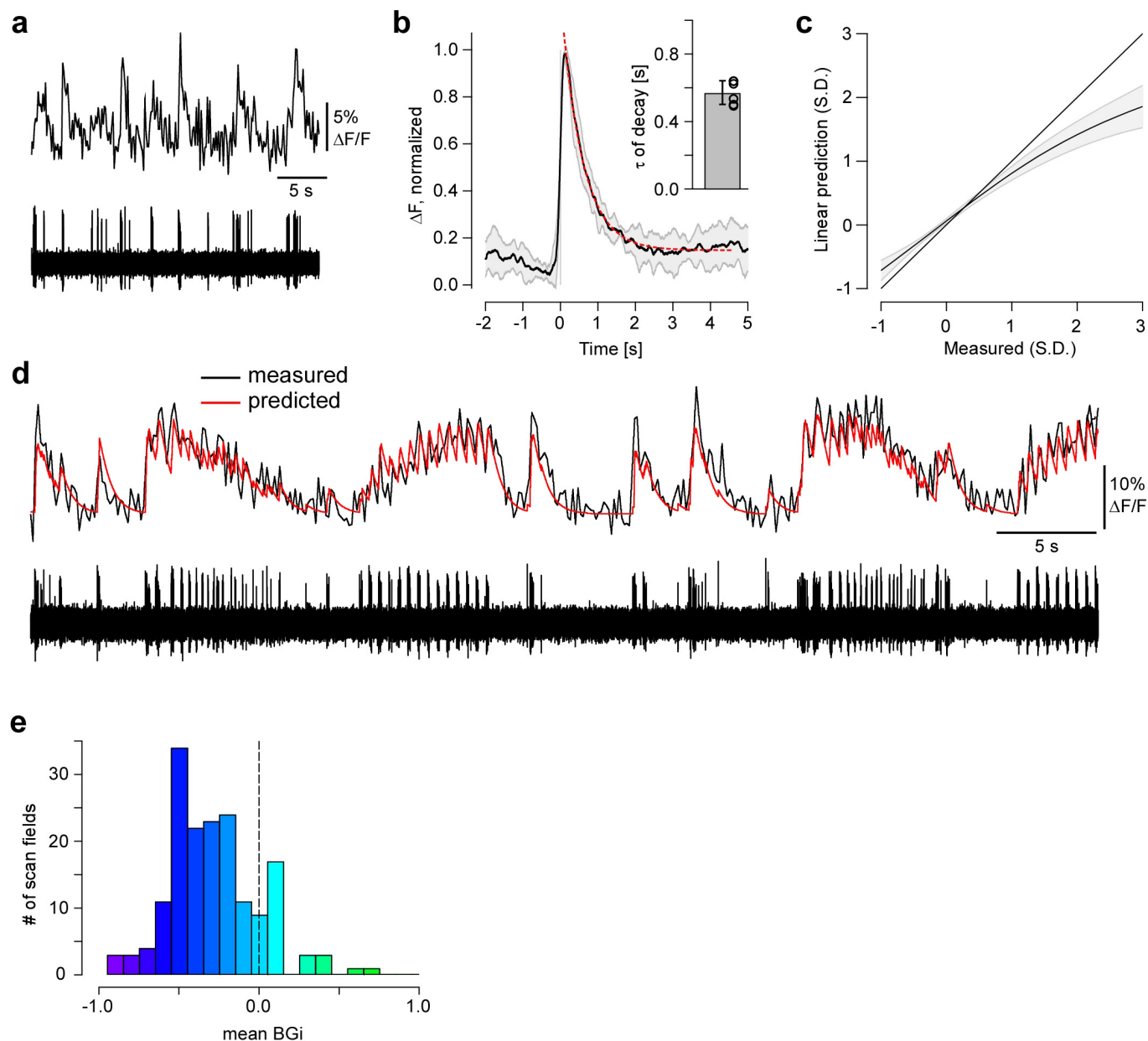
with the number of cells in a group (n_{Cells}), the median RF size ($A_{\text{RF,median}}$) within a group counting only cells that surpassed a RF quality criterion of 0.3, and the total scan area across all experiments (A_{Scan}). We corrected n_{Cells} for 29% cells discarded

by our quality criterion as well as an empirically estimated 8% of cells that did not yield a ROI in the first place due to weak or absent labelling. In addition, A_{Scan} was corrected for an empirically estimated 34.8% RF overhang (that is, where a cell's RF exceeds the scan field edge). This procedure yielded a CF of 2.0 ± 0.7 for most RGC groups (Gaussian fit; see Fig. 2e, right). However, differences between studies in approaches to measure RFs (for example, checkerboards vs. bars), in the assumptions used for RF fitting (for example, homogenous RFs best fitted by Gaussians), or in the methods to estimate dendritic arbor area can easily yield different absolute estimates of CF (see also Supplementary Table 1).

Estimation of IPL stratification profiles. To determine a cell's IPL stratification profile, we calculated dendritic density as described previously¹⁰ with spatial smoothing of $1 \mu\text{m}^3$. The resultant 3D density cloud was projected on the z axis to estimate the mean IPL depth profile. The relationship between the depth profiles and the two ChAT bands was estimated in independent experiments using mice that express tdTomato in cholinergic ACs ($\text{ChAT}^{\text{Cre}} \times \text{Ai9:tdTomato}$). We compared the IPL depth of the tdTomato-labelled dendritic plexi to the two SR101-labelled blood vessel planes that line the inner retina. We estimated the error to be $\sim 1.5 \mu\text{m}$ (s.d.), corresponding to 3–4% IPL depth ($n = 13$ measurements in 2 mice).

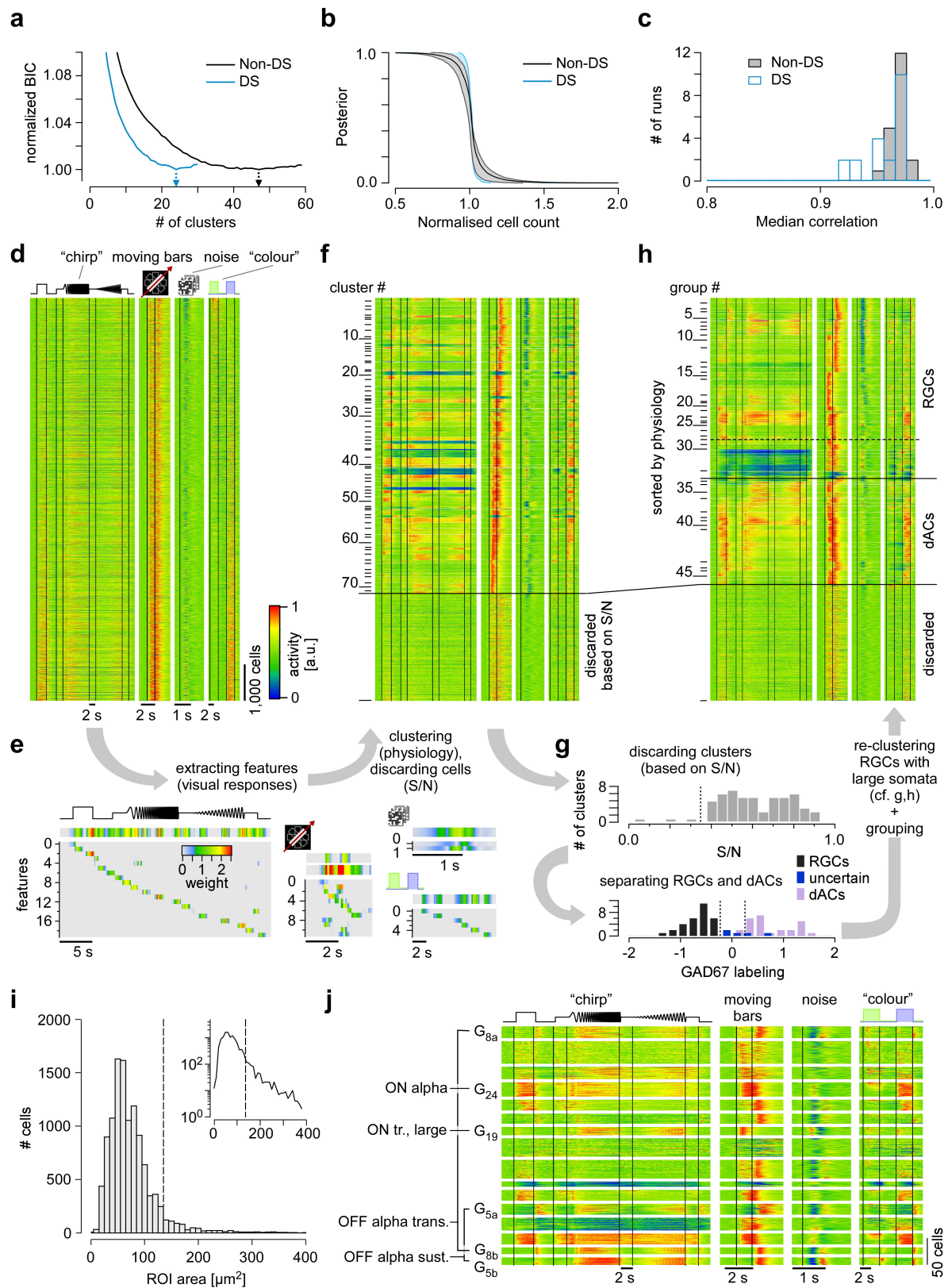
To relate each cell's IPL profile to functional groups we calculated the mean correlation coefficient between a cell's response to the chirp and moving bar stimuli and each group's mean response. The correlation coefficient ($-1 \dots 1$) for each pair was then multiplied with the cell's depth profile and a correlation-rank based weighting factor $W = 0.9^{\text{rank}-1}$. Thus, each individual recording yielded a complete two-dimensional map, with IPL depth on one axis and functional group on the other. Next, we averaged across the maps for those cells that passed our response quality criterion ($n_{\text{non-DS}} = 31/51$; $n_{\text{DS}} = 24/33$; see above). The resultant matrices were normalized in two steps: First, we divided each group's IPL depth profile by the mean depth profile of all included cells to eliminate any bias in sampling depth. Second, we divided each depth profile by its own maximum. This resulted in an automatic and unbiased estimate of dendritic stratification depth for all RGC groups (Fig. 5). Note that this automated approach is based on a relatively small sample of reconstructed cells and therefore can only provide an approximate prediction of stratification levels. This approach is invariant to differences in lateral dendritic field dimensions that may be associated with retinal position (for example, refs 10,23).

51. Rossi, J. *et al.* Melanocortin-4 receptors expressed by cholinergic neurons regulate energy balance and glucose homeostasis. *Cell Metab.* **13**, 195–204 (2011).
52. Breuninger, T., Puller, C., Haverkamp, S. & Euler, T. Chromatic bipolar cell pathways in the mouse retina. *J. Neurosci.* **31**, 6504–6517 (2011).
53. Ecker, A. S. S. *et al.* State dependence of noise correlations in macaque primary visual cortex. *Neuron* **82**, 235–248 (2014).
54. Zou, H., Hastie, T. & Tibshirani, R. Sparse principal component analysis. *J. Comput. Graph. Statist.* **15**, 265–286 (2006).
55. Fraley, C. & Raftery, A. Model-based clustering, discriminant analysis, and density estimation. *J. Am. Stat.* **97**, 611–631 (2002).
56. Ivanova, E., Hwang, G. S. & Pan, Z. H. Characterization of transgenic mouse lines expressing Cre recombinase in the retina. *Neuroscience* **165**, 233–243 (2010).



Extended Data Figure 1 | Linking electrophysiology and imaging data (related to Fig. 1). **a**, Simultaneously recorded RGC Ca^{2+} (top) and spiking (bottom) activity in response to binary spatial dense noise stimulation. **b**, Average Ca^{2+} event triggered by a single spike, averaged across $n = 6$ cells (shading indicates 1 s.d.); event decay was fitted (red) using a single exponential (for time constant τ , see inset, mean ± 1 s.d.) to yield an estimated impulse response. **c**, A linear prediction of Ca^{2+}

(calculated by convolution of the impulse response with binarized spike traces) was compared to measured values to estimate the mean nonlinearity. **d**, Ca^{2+} (top) and spiking (bottom) response to the full-field chirp stimulus (Methods) simultaneously recorded in an RGC (red trace, Ca^{2+} signal predicted from spiking response). **e**, Number of scan fields as a function of blue/green index (BGI, see Methods) averaged over all ROIs in each field (Fig. 1a).



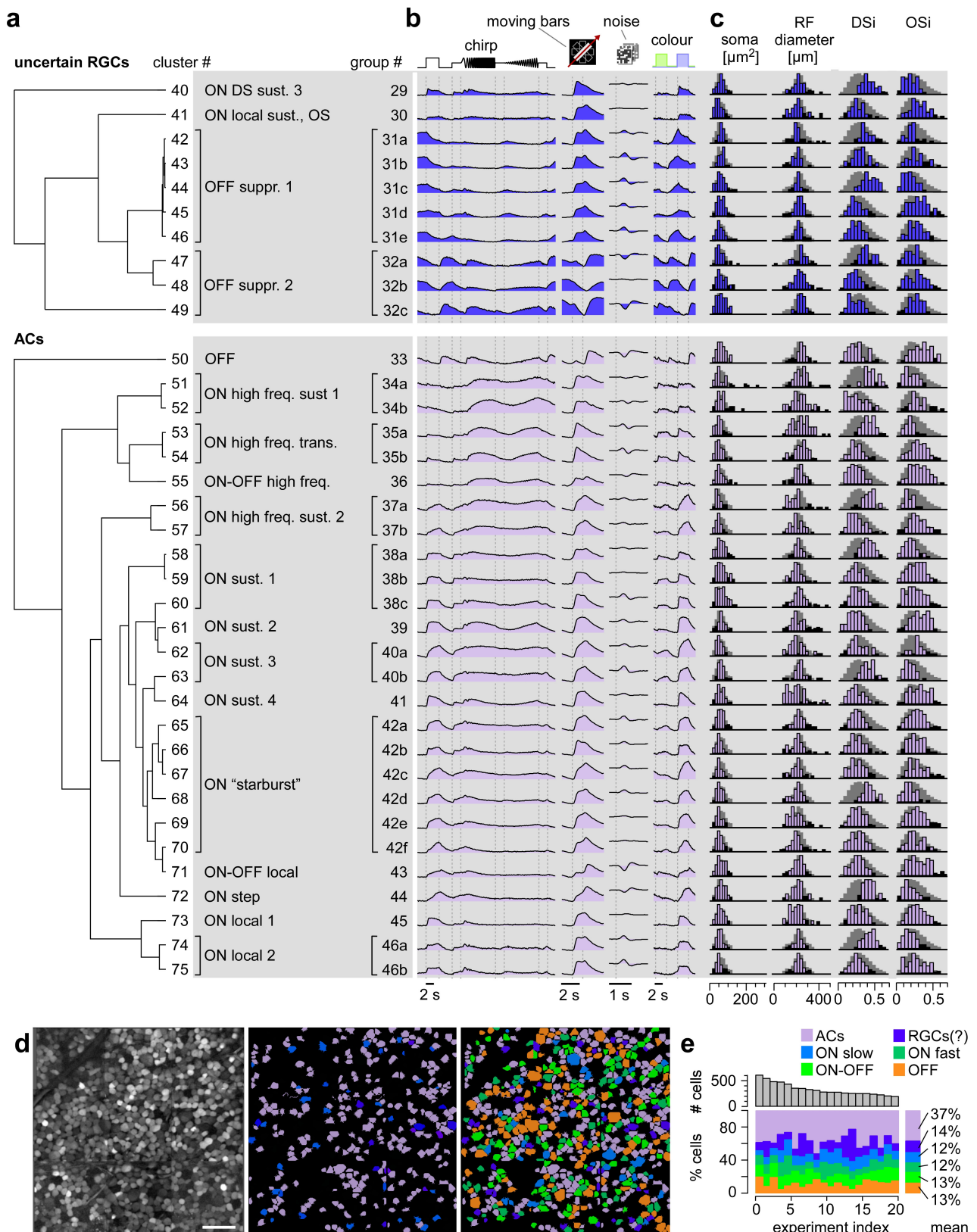
Extended Data Figure 2 | See next page for figure caption.

Extended Data Figure 2 | Clustering and grouping (related to Fig. 2).

a–c, Selection of cluster size and cluster quality/consistency analysis.

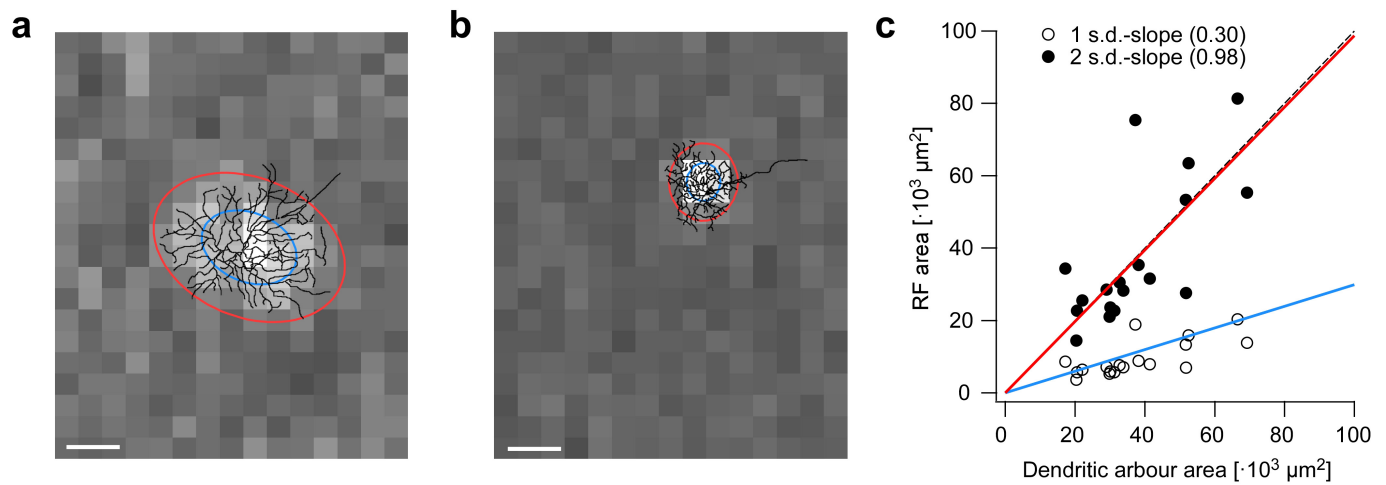
a, Normalized Bayesian information criterion (BIC) curves for non-DS (black) and DS (blue) cells. Arrows indicate the optimal numbers of clusters. **b**, Rank-ordered posterior probability curves indicating cluster quality. Curves were normalized for cluster size and averaged for non-DS (black) and DS (blue) clusters separately. Shaded area indicates 1 s.d. across clusters. **c**, Histogram of median correlation between the original clusters and clusters identified on 20 surrogate data sets, created by repeated subsampling of 90% of the original data set (bootstrapping); for each cluster, the best matching cluster from the original clustering was selected. **d**, Heat maps of Ca^{2+} responses to the four visual stimuli (see Fig. 1) of $n = 11,210$ cells from 50 retinas. Shown are raw data sorted by the response to the colour stimulus. Each line represents responses of a single cell with activity colour-coded such that warmer colours represent increased activity. **e**, Temporal features were extracted from the cells' light responses (Methods) and used for automatic clustering (**d** to **f**). **f**, Heat

maps showing clustered data ($n = 72$ clusters plus cells discarded based on signal-to-noise (S/N) ratio), with block height representing the number of included cells. **g**, Distributions of S/N (top) and GAD67 labelling (bottom) used to discard clusters and sort the remaining ones into retinal ganglion cells (RGCs), 'uncertain' RGCs and displaced amacrine cells (dACs). **h**, Heat maps showing $n = 46$ groups (divided into $n = 32$ RGC groups, including $n = 4$ 'uncertain' ones, and $n = 14$ dAC groups; sorted by response similarities) after re-clustering of large-soma cells (alpha cell post-processing, see panels **i**, **j**). **i**, Distribution of region of interest (ROI) area (as proxy for soma size) for all cells classified as RGCs and 'uncertain' (**g**). Inset, same distribution but on a log-scale. Dashed line marks threshold to separating large-soma cells (Methods). **j**, Results of re-clustering of large-soma cells (from **i**): heat maps show light-evoked Ca^{2+} responses to the four visual stimuli (see Fig. 1b). Clusters that resulted in new RGC groups are indicated; the remaining cells stayed with their original clusters.



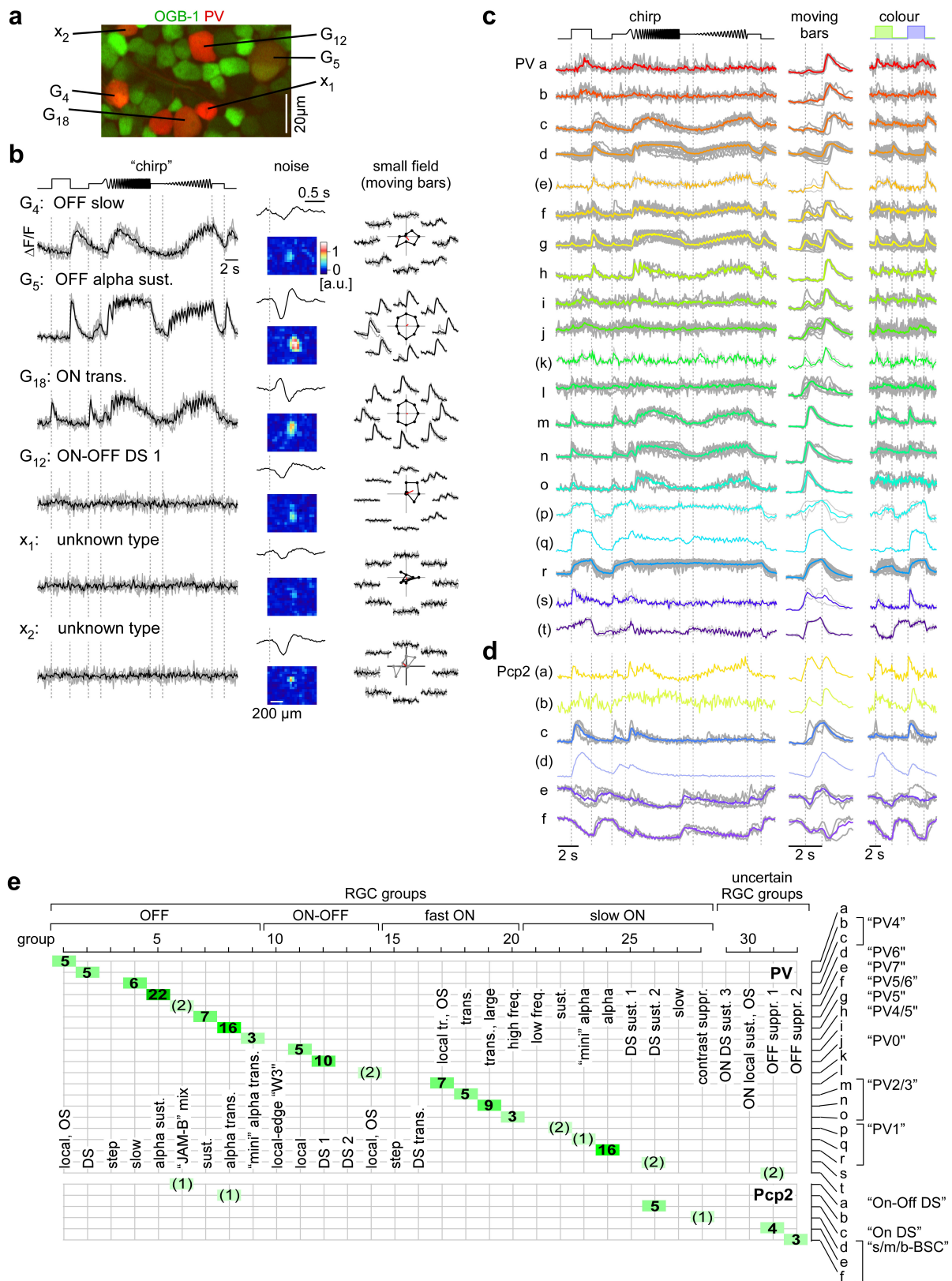
Extended Data Figure 3 | Group overview—functional groups classified as ‘uncertain’ RGCs and displaced amacrine cells (dACs) in the mouse retina (related to Fig. 2). **a**, Clusters organized according to hierarchical trees (dendrograms, see Methods) and grouped based on functional similarity (see main text for details), resulting in $n = 4$ ‘uncertain’ RGCs (top) and $n = 14$ dAC groups (bottom). **b**, Mean Ca^{2+} responses to the four stimuli (see Fig. 1b) for each cluster. **c**, Histograms of selected properties, from left to right: ROI (soma) area, receptive field (RF) diameter (2 s.d. from Gaussian fit; see Fig. 1b and Extended Data Fig. 4),

DS and OS indices (DSi and OSi, respectively, Methods). For details on each cluster, see also Supplementary Figures 1: 40–49 (‘uncertain’), and Supplementary Figures 1: 50–75 (dACs). **d**, Example experiment (left, from Fig. 1a); centre, dACs (lilac) and ‘uncertain’ RGCs (blue); right, colour-coded by broad categories, as in **e**. **e**, Total number of cells (top) and percentage of cells in sets of groups (bottom) per experiment (only experiments with ≥ 198 cells) illustrating consistency across experiments. Scale bar, 50 μm.



Extended Data Figure 4 | Relationship between RGC receptive field centres and their dendritic arbors (related to Fig. 2). **a, b**, Receptive field (RF) centre maps of a G₈ transient OFF alpha RGC (**a**) and a G₂ small-field RGC (**b**), with their reconstructed morphologies overlaid. 1- and 2-s.d.

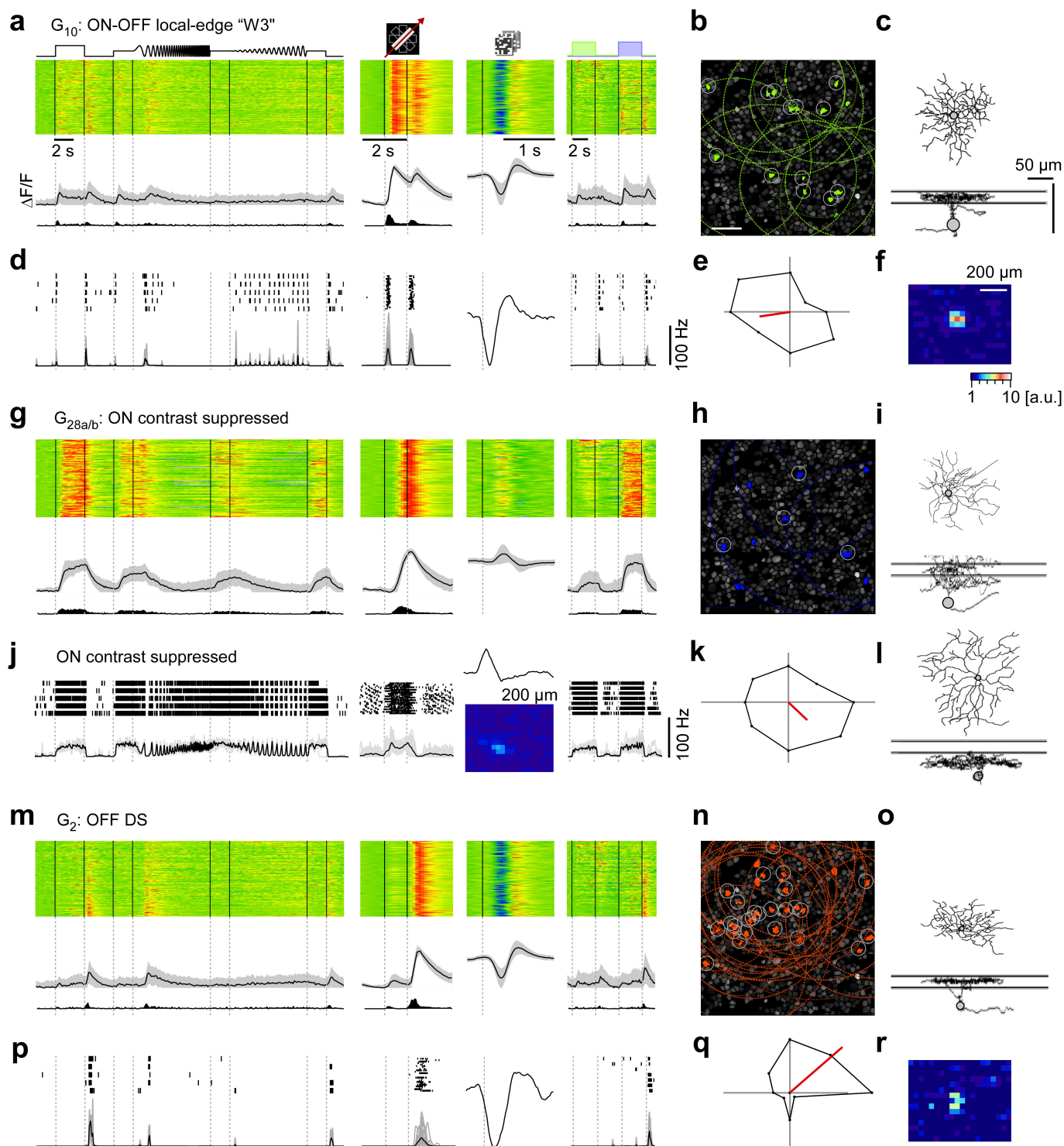
contours of RF centres fitted with 2D Gaussians are indicated by blue and red ovals, respectively. **c**, Area of RF centre fits from **a, b** as function of dendritic arbor area ($n = 18$ RGCs). Scale bars, 100 μm .



Extended Data Figure 5 | Mapping RGC groups onto genetic types—functional diversity of PV- and Pcp2-positive RGCs (related to Fig. 2).

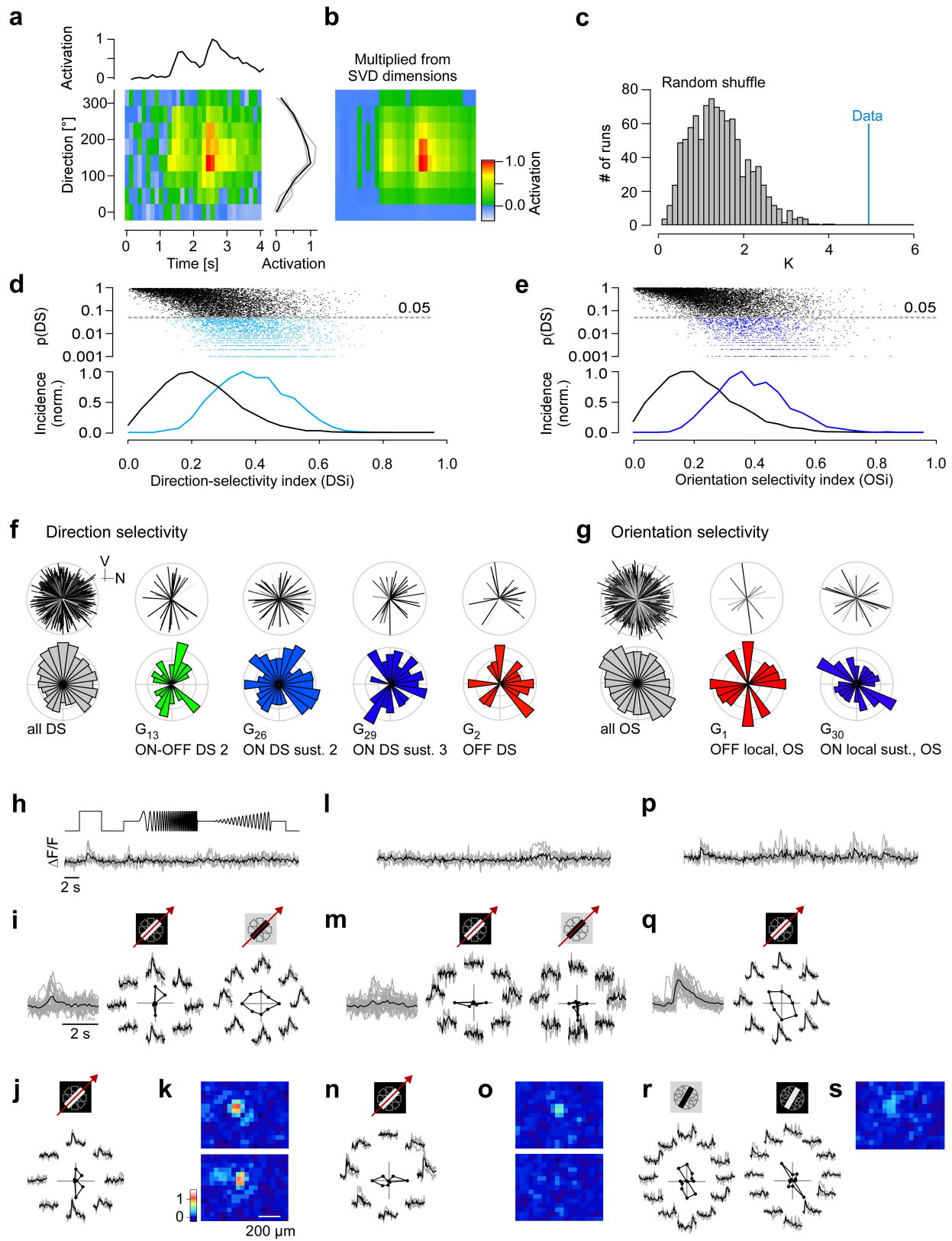
a, b, Diversity of PV-positive RGCs (red) in a PV:tdTomato mouse retina electroporated with OGB-1 (a, green). Ca^{2+} responses and receptive fields (**b**) from six PV-positive cells in exemplary field are shown (black, mean response, grey, single trials). The top four cells could be clearly matched to RGC groups (see Fig. 2), whereas the remaining two (x_1 , x_2) were discarded due to the lack of responses to both full-field and moving bar stimuli; note, however that both cells yielded a clear RF. **c**, Ca^{2+} responses

of functionally distinct PV-RGC groups (20 response types PV_{a-t}, thereof 14 with $n \geq 3$ cells). Traces colour-coded by group assignment (colours as in Fig. 2) represent mean responses, with individual cell responses in grey. **d**, Same for Pcp2-positive (six response types Pcp2_{a-f}, thereof three with $n \geq 3$ cells) RGC groups. **e**, Table illustrating the relationship between RGC groups (Fig. 2) and functional PV- and Pcp2-positive RGC types from (**c**, **d**). Numbers represent the total cell count of each allocation. Names in quotes (for example, "PV5") refer to the cell's original names (see PV (ref. 45) and Pcp2 studies (ref. 56)).



Extended Data Figure 6 | Examples of RGC groups. **a–c**, Functional 'fingerprint' of G_{10} RGCs, identified as local-edge-detector (W3) cells. Light-evoked Ca^{2+} responses of $n = 149$ cells: heat maps (top) illustrating individual responses, with response averages (with 1 s.d.) and firing rates estimated from Ca^{2+} signals (**a**; see Extended Data Fig. 1a–d) below. Ganglion cell layer (experiment from Fig. 1a) with G_{10} somata (green) and receptive fields (RFs, dotted) indicated (**b**). Grey circles mark cells with RFs that passed a quality criterion (Methods). Example morphology of a G_{10} cell filled after electrical single-cell recording (**c**). For a complete summary of the group's properties, see Supplementary Figure 2: 10.

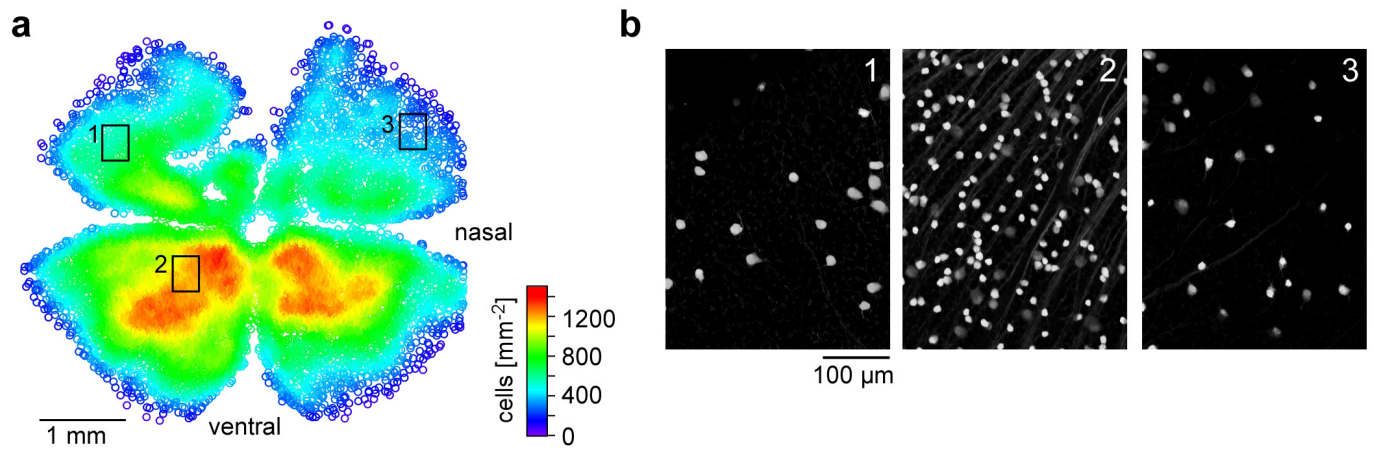
d–f, Electrical single-cell recording of a G_{10} cell: spiking responses as raster plots and mean spike rates for chirp, moving bar and blue/green stimuli as well as time kernel derived from noise stimulus (**d**), polar plot of responses to moving bar (**e**) and RF map (**f**). **g–i**, $G_{28a,b}$ ($n = 100$) contrast-suppressed ON RGCs with sample morphology (**i**; $G_{28a,b}$ cell dye-injected after Ca^{2+} imaging). **j–l**, Electrical single-cell recording of a contrast-suppressed ON RGC with different morphology (**l** vs. **i**). **m–r**, G_2 direction-selective OFF RGCs ($n = 162$) that stratify between the ChAT bands (**o**), as fingerprint (**m, n**) and exemplary electrical single-cell recording (**p–r**). Scale bars, 50 μm ; grey lines in **c, i, l, o**, ChAT bands.



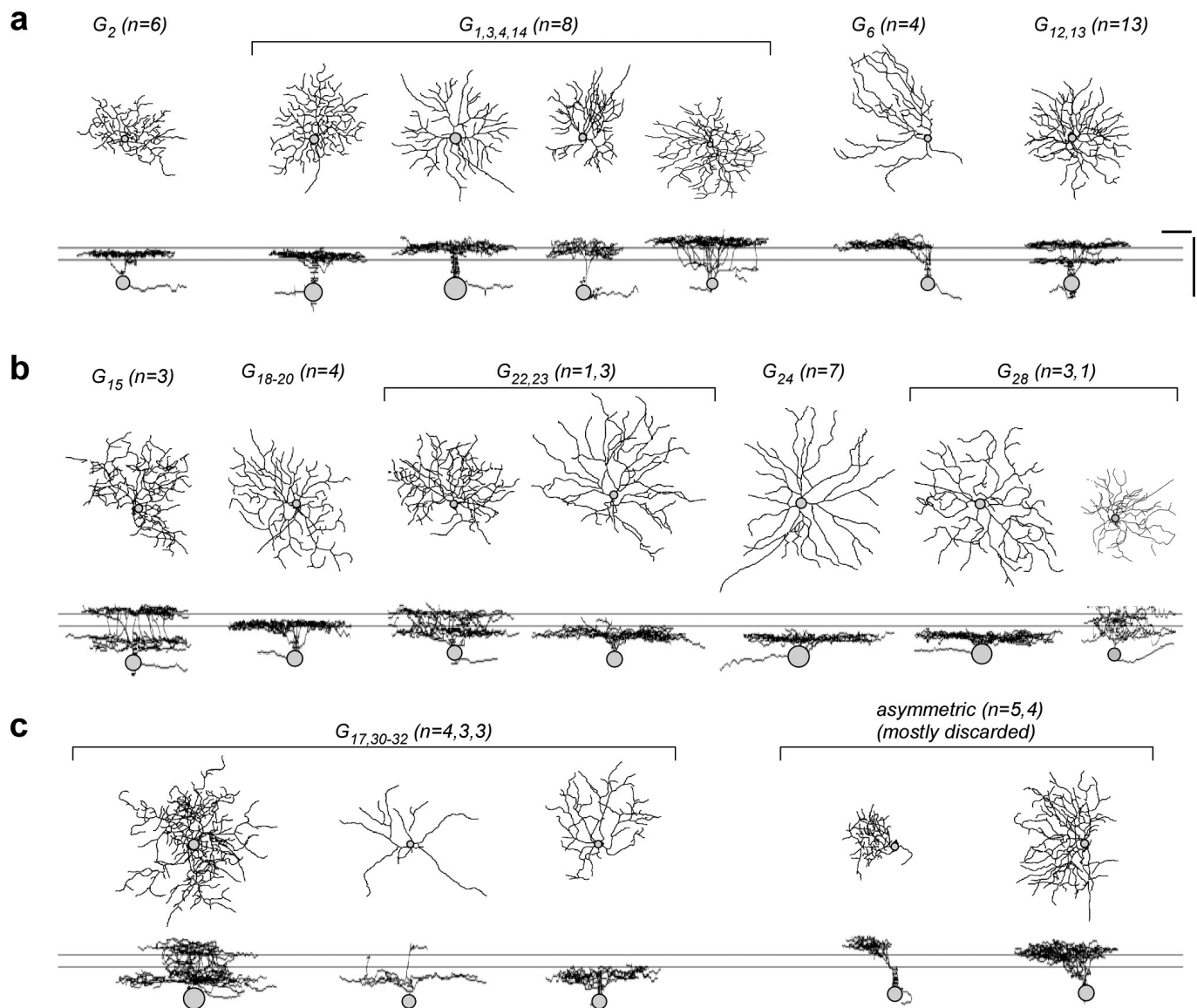
Extended Data Figure 7 | See next page for figure caption.

Extended Data Figure 7 | Direction and orientation selectivity (related to Fig. 4). **a**, Stimulus direction vs. time map for an exemplary direction-selective RGC with temporal (top) and directional (right) activation profiles shown; singular value decomposition (SVD) was used to estimate the time course and tuning function; individual stimulus repeats in grey, average in black. **b**, Reconstruction of direction vs. time map based on time course and tuning function of extracted by SVD. **c**, Statistical significance testing for direction selectivity (DS) or orientation selectivity (OS) was performed by projecting the direction/orientation profile on a single (for DS) or double (for OS) period cosine (blue) and the magnitude of the projection to the distribution of projections obtained by randomly permuting tuning angles from the original data (grey; bootstrapping). The P value is obtained by computing the percentile of the data (blue) in the bootstrap distribution (grey). **d**, **e**, P values for direction (**d**) and orientation (**e**) tuning as a function of the respective selectivity index (top, scatter plot; bottom, histogram; black, non-DS cells; light blue, DS cells; dark blue, OS cells). Note that tuning probability (p_{DS} , p_{OS}) only partially

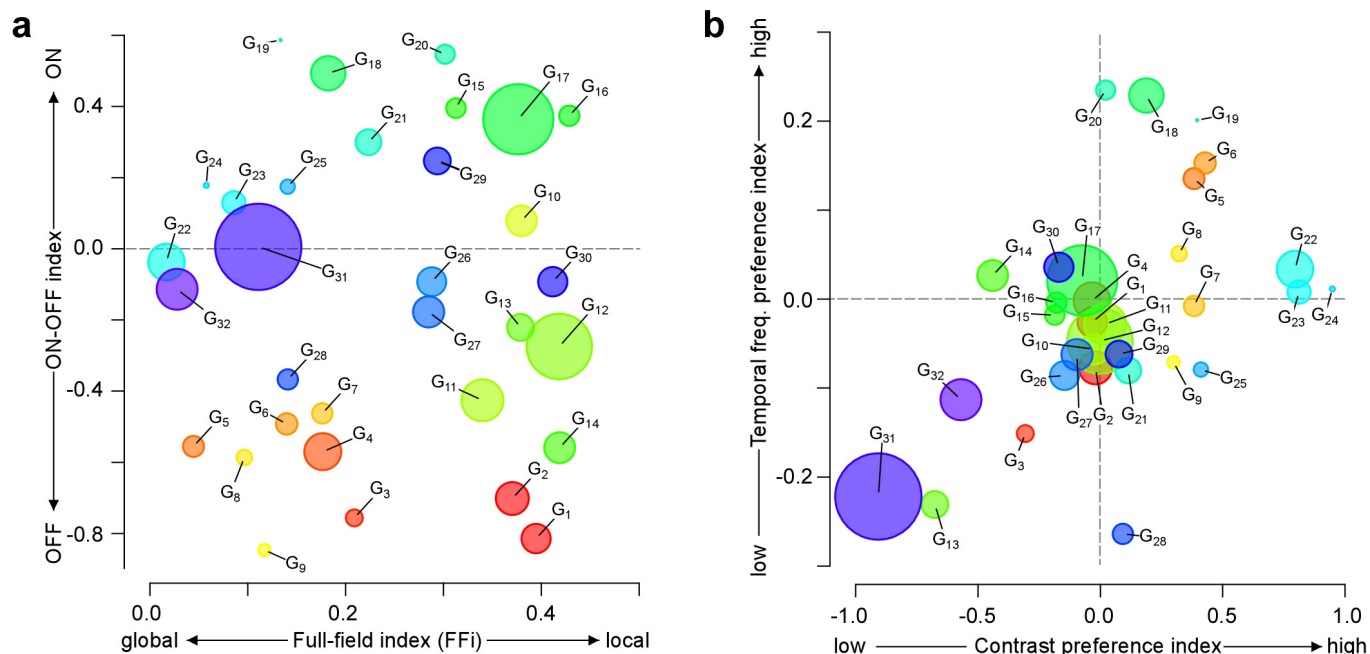
predicted tuning strength (DSi, OSi). **f**, Pairs of polar plots showing the distribution of preferred motion directions for all direction-selective (DS) RGCs together and for all DS RGC groups not shown in Fig. 4, (V, ventral; N, nasal direction; same group colour code as in Fig. 2). Top, plot of each pair: the cells' individual preferred directions, with line length representing DSi and line grey level p_{DS} (Methods). Bottom, plot of each pair: circular histogram of preferred direction. **g**, As for **f** but for orientation-selective (OS) RGCs. **h–s**, exemplary OS RGCs, illustrating the functional diversity within G_{17} (local ON trans. OS cells); none of them display strong full-field responses (**h**, **i**, **p**). A 'vertically tuned' ON OS cell (**i**, left) that shows little tuning to a dark moving bar (**i**, right; **j**, another example). Note the lobular structures bracketing the RF centre (coloured RF maps in **k**). **m–o**, Two examples for 'horizontally tuned' ON OS cells (**m**, **n**) with their respective RF maps (**o**). **p–s**, ON OS cell that shows weak tuning to bright moving bars (**q**) but strong OS to stationary bright and dark bars (**r**, left and right, respectively; Methods).



Extended Data Figure 8 | Retinal distribution of PV-positive cells in the $PV^{Cre} \times Ai9^{tdTomato}$ mouse line (related to Fig. 2). a, b, Density map (a) and magnified sample areas (b) illustrate PV-labelling anisotropy.



Extended Data Figure 9 | Mapping RGC groups to morphologies. a–c, Exemplary morphologies of RGCs filled after electrical recording or Ca^{2+} imaging and subsequently clustered/sorted into specific RGC groups or discarded (c, right) based on their light-response S/N. Scale bars, 50 μm .



Extended Data Figure 10 | RGC groups cover a basic feature space.

a, b, Relationship of four basic response indices of RGC groups. Disc area shows group size. Indices capture preference for stimulus polarity (ON-OFF index; Methods), for high vs. low temporal frequencies and contrasts (see below), as well as the full-field index (FFi; Methods), which reflects response preference for global (full-field chirp) versus local (moving bar)

stimulation. Contrast and frequency indices represent contrasts of feature activation ($(F_j - F_k) / (F_j + F_k)$) at respective time points during the full-field chirp stimulus, with $j = 12$, $k = 9$ for frequency, and $j = 17$, $k = 15$ for contrast. Before calculating ratios, feature activation (F) was normalized (0...1) by passing values through a cumulative normal distribution.

Divergent clonal selection dominates medulloblastoma at recurrence

A list of authors and affiliations appears at the end of the paper

The development of targeted anti-cancer therapies through the study of cancer genomes is intended to increase survival rates and decrease treatment-related toxicity. We treated a transposon-driven, functional genomic mouse model of medulloblastoma with ‘humanized’ *in vivo* therapy (microneurosurgical tumour resection followed by multi-fractionated, image-guided radiotherapy). Genetic events in recurrent murine medulloblastoma exhibit a very poor overlap with those in matched murine diagnostic samples (<5%). Whole-genome sequencing of 33 pairs of human diagnostic and post-therapy medulloblastomas demonstrated substantial genetic divergence of the dominant clone after therapy (<12% diagnostic events were retained at recurrence). In both mice and humans, the dominant clone at recurrence arose through clonal selection of a pre-existing minor clone present at diagnosis. Targeted therapy is unlikely to be effective in the absence of the target, therefore our results offer a simple, proximal, and remediable explanation for the failure of prior clinical trials of targeted therapy.

Extensive efforts to understand the molecular underpinnings of medulloblastoma^{1–7} are driven by the desire to develop rational, targeted therapies that will increase survival rates, and diminish the considerable complications of radiotherapy and cytotoxic chemotherapy⁸. The development of targeted therapy for medulloblastoma has been hampered by the relative paucity of somatic single nucleotide variants (SNV), the low tumour incidence compared to adult epithelial malignancies, and the existence of four distinct molecular subgroups (Shh, Wnt, Group 3, and Group 4)^{9,10}. The common practice in paediatric oncology is for novel agents to be tested in phase I and/or phase II trials that enroll children previously treated with radiotherapy and cytotoxic chemotherapy. The majority of basic and translational research on the biology of medulloblastoma makes use of samples or models of medulloblastoma that have not been exposed to prior anti-tumour therapies. There are very few genomic studies on recurrent medulloblastoma, as recurrent disease is nearly universally fatal, and surgery at the time of recurrence is associated with significant morbidity and discomfort¹¹. The current clinical pathway in which new agents are tested at recurrence is therefore based on the unsubstantiated premise that the recurrent tumour is biologically and genetically highly similar to the tumour at diagnosis, and therefore well represented by tumour models derived from pre-treatment tissue samples.

Recent genomic approaches in liquid cancers (frequently re-biopsied) have suggested that the tumour genome at the time of recurrence is divergent from the genome at diagnosis^{12–17}, as seen in some solid cancers^{18–20}. Critical and careful examination of human cancer xenografts clearly demonstrates clonal evolution^{21–23}, even in the absence of therapy. Almost all medulloblastoma research to evaluate novel agents has been carried out with cell lines or xenografts derived from naive biopsies, or mouse models in which the experimental therapy is provided at diagnosis (not after standard therapy). Successful phase I or phase II trials of novel agents are uncommon in paediatric oncology, particularly for targeted agents, and almost completely non-existent for medulloblastoma. We hypothesized that recurrent medulloblastoma is highly genetically divergent from patient matched pre-therapy disease, current experimental models fail to model recurrent disease, and that genetic divergence with loss of targets at recurrence could account for the lack of success seen in clinical trials.

A mouse model of recurrent Shh medulloblastoma

To develop an *in vivo*, functional genomic, ‘humanized’ mouse model of recurrent medulloblastoma we studied *Ptch*^{+/-} mice that have transposition of the *Sleeping Beauty* (SB) transposon in the *Math1* compartment of the developing cerebellum (*Ptch*^{+/-}/*Math1*-SB11/*T2Onc* or *T2Onc2*). These mice have a high penetrance and short latency to develop metastatic sonic hedgehog (Shh) medulloblastoma^{7,24}. To accurately model recurrent medulloblastoma, we performed subtotal tumour removal for 38 *Ptch*^{+/-}/*Math1*-SB11/*T2Onc* or *T2Onc2* mice. (Extended Data Fig. 1a). Standard therapy for children with metastatic medulloblastoma includes multi-fractionated image guided craniospinal irradiation (CSI) to 36 Gy over four weeks. After surgery, mice received 18 fractions (2 Gy each) of CSI over four weeks. To selectively target the central nervous system (CNS) and to spare targeting non-CNS tissues, we used two-dimensional (2D) fluoroscopic images (Extended Data Fig. 1b) and three-dimensional (3D) volumetric cone-beam CT (computed tomography) images (Fig. 1a). After completion of therapy, mice were monitored for tumour recurrence. The combination of microsurgical resection followed by image guided fractionated CSI allows us to accurately mimic the therapy given to children with medulloblastoma. Using an intent-to-treat analysis, mice treated with surgery and CSI have an increased medulloblastoma-free survival compared to untreated controls (Fig. 1b), median survival is 118 days for the treated group, and 5 days for the control group. However, 11/18 (61%) of treated mice developed local and/or metastatic relapse (Extended Data Fig. 1c).

Genetic divergence in recurrent mouse medulloblastoma

Massively parallel sequencing of transposon insertions analysed using a significant recurrence method identified 23 gCISs (genetic Common Insertion Sites) (Supplementary Table 1a) in 11 primary tumours, and 40 gCISs from the local and metastatic recurrences^{7,25}. gCISs in the diagnostic samples are extremely different from those at recurrence (Fig. 1c), with only the known medulloblastoma tumour suppressor gene *CBP* (also known as *CREBBP*) found across all compartments, and only *Trp53*, *Arid1b*, and *Tcf4* identified in both recurrent compartments (Supplementary Table 1a). We also developed a complementary computational method based on the concept that tumour cell doublings

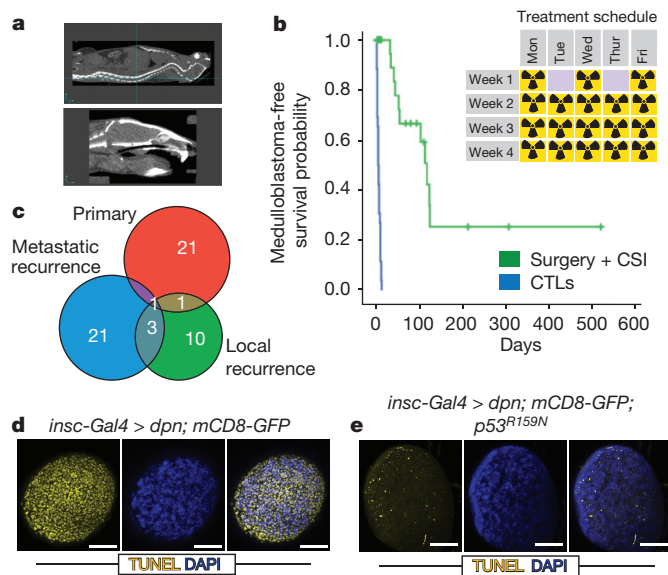


Figure 1 | A novel functional genomic mouse model of recurrent Shh medulloblastoma using microneurosurgical resection and computed-tomography-guided multi-fractionated craniospinal radiotherapy. **a**, *Ptch*^{+/-}/*Math1*-*SB11*/*T2Onc* mice with medulloblastoma underwent subtotal tumour removal ($n = 38$) and received multi-fractionated CSI post-operatively. Radiation was delivered under computed tomography (CT) guidance. **b**, Microneurosurgery and CSI strikingly improves tumour-free survival as compared to untreated controls ($P = 0.0001$, log-rank test, $n = 64$). Inset schematic indicates the fractionation schedule. **c**, Venn diagrams demonstrate the paucity of overlap in the gCISs between primary tumours and their recurrences. **d**, *Drosophila* brain tumours harbouring wild-type P53 displayed massive apoptosis in response to 40 Gy irradiation. **e**, Dominant negative P53 (*p53*^{R159N}) essentially abrogated the radiation-dependent cell death. Scale bar, 50 μ m.

will produce a characteristic pattern of *Sleeping Beauty* insertion site frequency, with driver-initiating insertions being most frequent. Results from this strategy predict driver events that show a cancer pathway enrichment, replicate the lack of overlap between events important for primary tumorigenesis versus those at recurrence, and identify loss of function of *Trp53* as a key event in the pathogenesis of recurrence (Supplementary Table 1b and Supplementary Note).

Clonal transposon insertions in *Trp53*, *Arid1b* and *Tcf4* are distributed across the coding region of genes and are predicted to result in loss of function (Extended Data Fig. 1e)^{26,27}. End-point PCR for *Trp53* insertions shows that they are highly clonal in the recurrences, and either absent or present in a subclone of the matched primary tumours (Extended Data Fig. 1f). Similarly, driver events predicted by our complementary computational method that are clonally prominent at diagnosis are reduced at recurrence, while events prominent at recurrence are not seen at diagnosis (Supplementary Table 1c). We conclude that medulloblastoma recurrences have undergone substantial genetic divergence, possibly due to clonal selection secondary to surgery and radiation.

Humans with germline mutations in *TP53* develop Li-Fraumeni syndrome and are at increased risk of developing medulloblastoma. Somatic mutations of *TP53* are also found in sporadic human medulloblastomas²⁸. Human Shh medulloblastomas with mutations in *TP53* are almost always fatal, as they fail to respond to current therapies including radiation²⁸. Similarly, we find that animals whose recurrence contains a clonally selected *Trp53* insertion (gCIS), exhibit a worse prognosis (Extended Data Fig. 1g). In our model, driver *Trp53* insertions are not clonal at diagnosis, but become clonal at recurrence (Supplementary Table 1c). We observe decreased expression of the *TP53* transcriptional target p21 in recurrent medulloblastomas with gCIS in *Trp53*, indicating loss of function (Extended Data Fig. 2a).

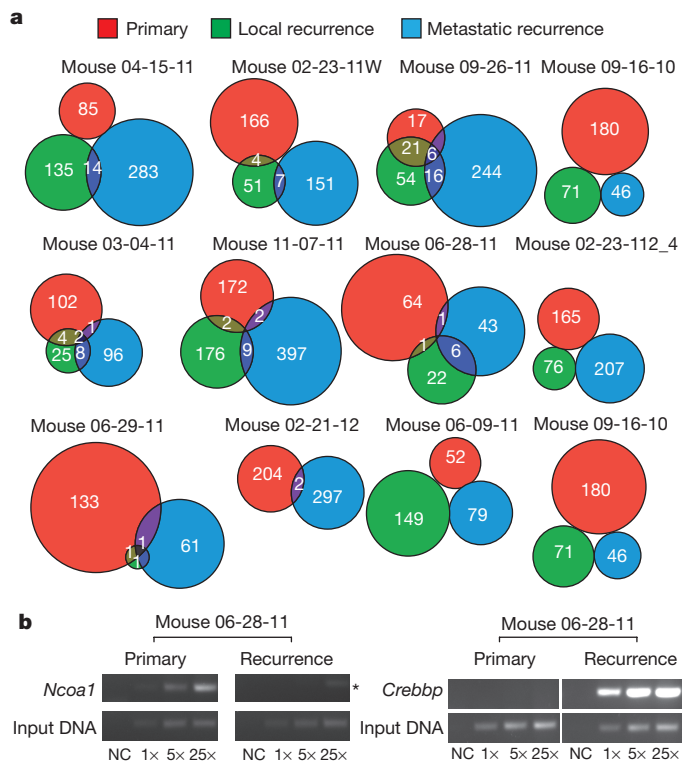


Figure 2 | Paucity of shared genetic events between therapy-naive and recurrent tumours in individual mice treated with microneurosurgery and CT-guided multifractionated craniospinal radiation. **a**, Venn diagrams demonstrate the paucity of clonal insertions shared between therapy-naive primary tumours and their matched local and metastatic recurrences. Matched recurrences share only very few clonal transposon insertions with the paired primary tumour. **b**, End-point PCR demonstrates examples of highly clonal insertions that are restricted to the untreated primary (*Ncoa1*) and the recurrence (*Crebbp*), (asterisk indicates non-specific amplification). Three levels of input DNA were used for each sample 1 \times , 5 \times and 25 \times ; NC, negative control.

To further demonstrate the consequence of defective p53 signalling on the therapeutic outcome after radiation *in vivo*, we studied a *Drosophila* neuronal brain tumour model, in which we expressed the oncogene *dpm* using *insc*-Gal4 in the neural stem cell lineage²⁹. Treating the third instar larvae bearing brain tumours with 40 Gy irradiation resulted in widespread apoptosis in the brain lobes (Fig. 1d, e). Strikingly, expressing a dominant negative form of the *Drosophila* p53, *p53*^{R159N} (ref. 30), essentially abrogated this radiation-induced cell death response (Fig. 1d, e), while moderately increasing the level of mitosis (Extended Data Fig. 1h).

Examination of the transposon insertion patterns in each compartment by animal (Fig. 2a) confirms the paucity of overlap in clonal genetic events between matched diagnostic and recurrent tumours. Although some insertions that assume clonal dominance at recurrence probably arise due to selection, others are true *de novo* insertions not found in the primary tumour (Fig. 2b and Extended Data Fig. 2b). Many other clonal insertions in the recurrence are indeed found in very restricted subclones of their matching primary tumour (Extended Data Fig. 2c–e). We used pathway analysis to compare the driver events present in the diagnostic samples versus the local recurrent samples. Initiating driver insertions were enriched in the Shh signalling pathway at diagnosis, but not at recurrence. Conversely, the locally recurrent samples had over-representation of insertions in the *TP53* pathway, as well as in genes important in the DNA damage response that were not seen at diagnosis (Supplementary Table 1d). We conclude that in our mouse model of recurrent Shh medulloblastoma, the recurrent tumour has undergone a major change in its

biology through a process of clonal selection, and that most targets identified in the diagnostic sample are unlikely to be present in the dominant clone at the time of recurrence.

Genetic divergence in recurrent human medulloblastoma

To validate our mouse model of recurrent medulloblastoma, we collected therapy-naïve primary human medulloblastoma samples and patient-matched recurrent tumours (after radiation and chemotherapy, or radiation alone). Therapy-naïve and recurrent medulloblastoma tumour pairs with ($n = 15$) and without ($n = 18$) matched germline DNA were profiled by whole-genome sequencing (WGS). An additional 10 recurrent tumours with matching germline also underwent WGS (no pre-therapy biopsy available). Three additional therapy-naïve tumours with matching recurrences were also profiled from formalin fixed paraformaldehyde embedded (FFPE) material using whole-exome sequencing (WES), for a total of 46 patient samples (Supplementary Table 2a, b, g, h).

We performed integrative analysis of somatic SNVs, copy number aberrations (CNAs), loss of heterozygosity (LOH), short indels, and large-scale structural rearrangements from our WGS data. Our results demonstrate striking genetic divergence between therapy-naïve medulloblastoma and patient-matched recurrent medulloblastoma across all subgroups (Fig. 3a–c and Supplementary Information). Strikingly, in 13/15 patients the somatic mutational burden increased by an average of fivefold at recurrence (P value = 2.7×10^{-4} ; Fig. 3a). Although two exceptions (MB-REC-15/16) have more somatic SNVs than the average therapy-naïve tumour, their matched recurrences are similar to other recurrent medulloblastomas. In every case, only a minority of genetic events is shared by therapy-naïve and patient-matched recurrent tumours (Supplementary Information and Supplementary Table 2c, d). Similar to our mouse model, on average, only 11.8% of human somatic SNVs and indels were present in both the diagnostic and recurrent samples, demonstrating a substantial genetic divergence at recurrence.

We classified mutations based on their allelic frequencies (Methods; ref. 31) and noted a significant increase in the proportion of clonal mutations post-therapy (1.9-fold increase; P value = 8.7×10^{-3} ; Extended Data Fig. 3, Supplementary Information and Supplementary Table 2j), an observation consistent with therapy-induced selection eliminating low-frequency lineages in the primary tumour. Specifically, the majority (60.5%) of damaging clonal mutations present in primary tumours consistently decrease in abundance post-therapy and either become subclonal (25.9%), or disappear completely (34.6%). Only 25% of patients retain the full set of clonal SNVs post-therapy, with most patients having no retention (41.6%) or partial retention (33.3%) of clonality. Strikingly, damaging clonal mutations post-relapse outnumber the clonal events retained from the primary tumour by fivefold, indicating the importance of profiling this compartment.

In cases without germline controls, we enriched for somatic variants by considering events restricted to either the therapy-naïve or the recurrent tumour, and found the same trend of increased mutational burden at recurrence (Fig. 3a). Comparison of mutational spectra between the therapy-naïve and recurrent compartments reveals four main SNV signatures whose relative proportions change significantly at relapse (Extended Data Fig. 4a, b). Consistent with observations in AML¹⁴, recurrences had a significantly increased proportion of transversions, possibly due to DNA damage induced by therapy (Extended Data Fig. 4c–e).

Structural aberrations are also observed at a higher frequency at recurrence (Fig. 3c; Supplementary Table 2e, f and Supplementary Information), with a subset of aberrations identified at diagnosis no longer observed at recurrence. For instance, the diagnostic sample of MB-REC-14 harbours a *TERT* amplification absent at recurrence (Extended Data Fig. 5a). Discordance across time makes *TERT*³² a less desirable therapeutic target in this patient at recurrence. Conversely,

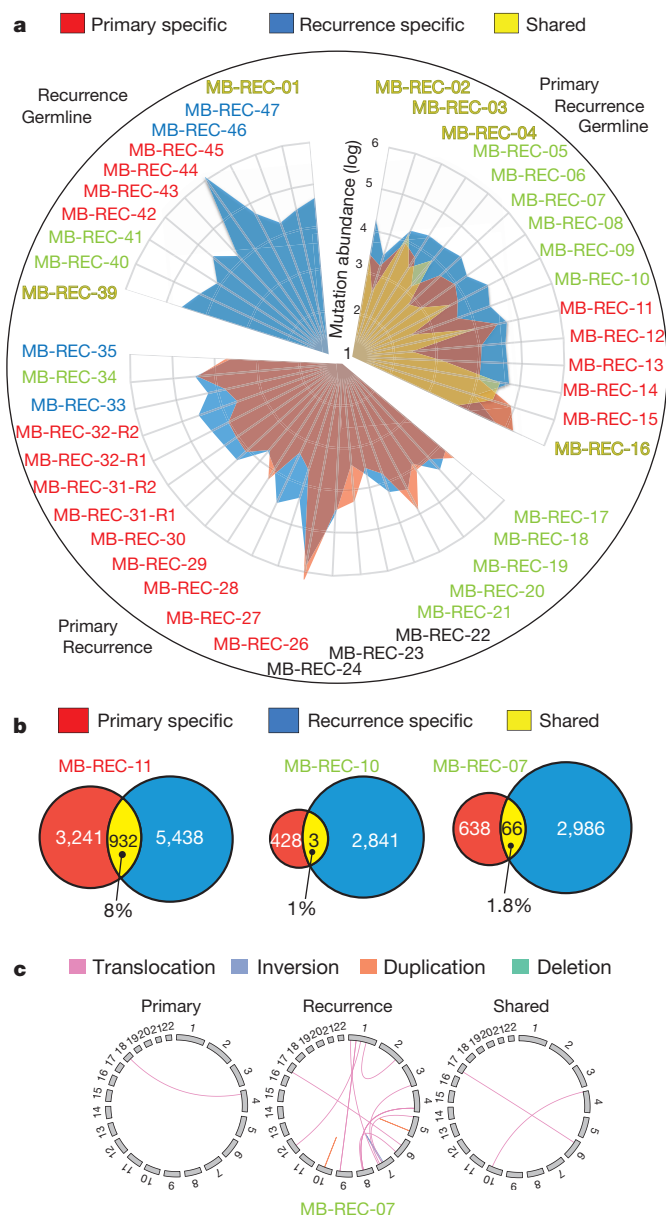


Figure 3 | Major genetic divergence of human untreated medulloblastoma and patient-matched recurrences determined by whole-genome sequencing. a, Somatic mutation burden in 45 tumours (43 patients) was increased fivefold in matched post-treatment (blue) versus therapy-naïve (red) tumours (Student's t -test; P value = 2.7×10^{-4}). On average, 11.8% of mutations are shared somatic events ($n = 15$ cases with germline). Hypermutated samples stand out by two orders of magnitude (MB-REC-26/44). Patient subgroup is indicated by the label (blue, Wnt; red, Shh; yellow, Group 3; green, Group 4; black, undetermined). b, Venn diagrams of three representative patients reveal a minimal overlap in genetic events between therapy-naïve (red) and recurrent (blue) tumours. c, Circos plot in a representative patient illustrates compartment-specific somatic structural variations.

MB-REC-09 demonstrates chromothripsis involving the *MYC* locus only at recurrence³³ (P value = 3.97×10^{-7} , Extended Data Fig. 5b).

Two-thirds of patients have deleterious events in at least one gene for which an anti-neoplastic drug interaction has been defined³⁴ ($n = 9$ of 15; Extended Data Fig. 5c and Supplementary Table 2i). The current assumption that all putative drug targets present at diagnosis are retained post-therapy is only valid in 4 of these patients (44.4%), though in one of these cases, we see different mutations in *SMO* before and after therapy, indicating convergent evolution. Of the other five

patients, two have a complete switch in actionable targets from the naive to the post-therapy samples, and three patients gain actionable targets post-therapy (Extended Data Fig. 5c). Cumulatively, our data demonstrate that putative drug targets discovered in therapy-naïve tumours are not, as previously assumed, representative of the targets present at recurrence. A number of targetable events are present in the recurrent tumour as subclonal mutations, indicating that a combination therapy approach may be necessary to minimize evolution of resistance³⁵.

Clonal selection drives recurrence

Patient MB-REC-12 harbours a clinically compelling example of a homozygous *PTCH1* driver mutation that is clonally dominant in the primary tumour and completely eradicated by therapy (Supplementary Table 2k). The nature of the *PTCH1*^{-/-} driver (heterozygous mutation followed by chr9q LOH) reveals that the recurrent tumour is derived from an ancestral lineage with wild-type chr9q heterozygosity (Fig. 4a, b). Thus, cells derived from the ancestral clone remained present at low frequency in the primary tumour, and a sub-lineage of these cells driven by *CDKN2A* and *CDKN2B*^{-/-} loss successfully reconstituted the tumour post-therapy. The shift in driver from *PTCH1* does not change the subgroup affiliation of the tumour, as this and other Shh patients retain a Shh-like transcriptional signature (Supplementary Table 2l). This complete switch in 'trunk' driver mutations post-therapy highlights the evolutionary plasticity of medulloblastoma, and imparts a cautionary note to the therapeutic strategy of targeting trunk mutations.

We used EXPANDS³⁶ to computationally model the clonal diversity of the primary and recurrent tumours and globally assess clonal dynamics as a function of therapy. Leveraging genome-wide mutation and copy number data, EXPANDS infers a branched evolution pattern in 14 primary and recurrent tumour samples with matched patient germline (Fig. 4c; Extended Data Fig. 6a and Supplementary Table 3a). In the majority of patients (8/14), all clones in the recurrent tumour arise from a single lineage in the primary tumour (Fig. 4c and Extended Data Fig. 6a). In the remaining 6 patients, we see a more intermediate (MB-REC-02/07/16) or high (MB-REC-03/13/14) phylogenetic similarity to the primary tumour. Most patients have an increased number of clonal populations post therapy (71.4%), and Group 4 tumours stood out as significantly more heterogeneous (diversity measured by the Shannon Index³⁷; *P* value = 0.029, Extended Data Fig. 6b and Supplementary Table 3b).

To more precisely delineate the extent of clonal selection in recurrent human medulloblastoma, we performed ultra-deep sequencing of 192 patient-specific SNVs (*n* = 20 patients; Supplementary Table 3c) and compared their cellular prevalence pre- and post-therapy using PyClone³⁸. PyClone identifies clusters of somatic SNVs that co-occur in the same lineage, and determines their cellular prevalence (the proportion of tumour cells harbouring the SNV). We observed three types of patterns consistent with the findings from EXPANDS: (1) subclonal lineages in the primary tumour expanding to become clonal at relapse, (2) clonal lineages in the primary tumour that become subclonal at relapse, and (3) lineages in the primary tumour that retain the same cellular prevalence upon tumour relapse (Fig. 4d and Extended Data Fig. 6c). In each and every tumour we observe a significant incidence of novel mutations restricted to the dominant clone at the time of recurrence (Extended Data Fig. 7a). On the basis of the results from our mouse model, we hypothesized that some of these events might exist as rare subclones at diagnosis. To determine the sensitivity to detect rare (<5%) sub-clonal events, we focused on clonal SNVs in the recurrences that had at least one read of support in the primary sample and confirmed the presence of subclonal events in the primary tumour present at frequencies as low as 2/10,000 (Extended Data Fig. 7b–d and Extended Data Fig. 8). Overall, we find evidence for significant expansion of clones initially present at <5% in the therapy-naïve tumour in 16/20 patients, indicating that clonal selection is

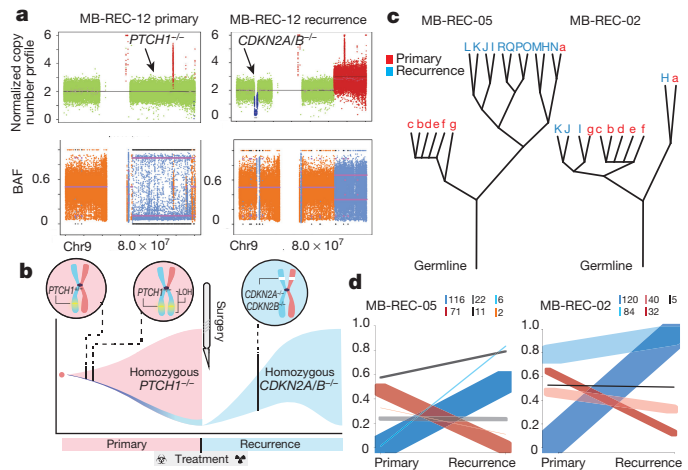


Figure 4 | Genetic divergence of recurrent medulloblastoma is driven by clonal selection. **a**, Copy-neutral LOH *PTCH1*^{-/-} driver status reverts to wild type post-therapy in medulloblastoma-REC-12, with homozygous *CDKN2A/B* loss. **b**, The evolutionary progression of medulloblastoma-REC-12 is illustrated by (pink) *PTCH1*^{+/-} lineage expansion, copy-neutral LOH, clonal eradication during treatment, and (blue) subsequent expansion of an ancestral clone with *CDKN2A/B*^{-/-}. **c**, Phylogenetic relationships between primary (red) and recurrent (blue) tumours show that recurrences often represent a single rather than multiple primary tumour lineages (for example, medulloblastoma-REC-05/12 compared with medulloblastoma-REC-02). **d**, Ultra-deep sequencing shows post-treatment expansion of low-frequency or *de novo* primary clones (blue), and eradication/reduction of therapy-sensitive lineages (red). Inset box indicates number of mutations per cluster.

commonly observed after therapy for medulloblastoma (Extended Data Fig. 7a).

Convergent biological pathways at relapse

We observe recurrently mutated genes and pathways restricted to the recurrent compartment (Extended Data Fig. 9 and Supplementary Table 2q). Similar to data from our mouse model, genetic events in *TP53* pathway genes (*n* = 12/23, 52.2%; KEGG04115) or the actual *TP53* gene (*n* = 6/23, 26.1%) are frequent in the human recurrences, predominantly in Shh medulloblastoma (Fig. 5). The deleterious effects of *TP53* on recurrence of Shh medulloblastoma are known, and we conclude that in both humans and mice, *TP53* pathway alterations constitute a common feature observed at recurrence of Shh medulloblastoma^{28,39}.

Additionally, recurrent damaging mutations in *DYNC1H1* are restricted to the post-therapy compartment in 16% of human Shh tumours (Fig. 5 and Supplementary Table 2q), and mutations in this gene are mutually exclusive with mutations in *TP53*. Therapy-naïve Shh medulloblastoma with deletion of one copy of the *DYNC1H1* locus (chr14q) have a poor prognosis, but chr14q loss has no prognostic impact for non-Shh medulloblastoma (Fig. 5b, c; Extended Data Fig. 10a–e and Supplementary Table 4a–d). Notably, recurrence rates were higher in patients with chr14q loss (50%, 7/14) as compared to balanced chr14q (30%, 7/23). Furthermore, cases with chr14q loss were mutually exclusive with *DYNC1H1* mutations, and show decreased *DYNC1H1* expression (Fig. 5 and Extended Data Fig. 10a, c). Cumulatively, 6/15 (40%) Shh medulloblastoma recurrences functionally lose one *DYNC1H1* allele. Taken together, we observed *TP53* gene and pathway mutations, *DYNC1H1* mutations, or chr14q losses in 79% of recurrent Shh medulloblastoma.

These genomically defined pathway signatures were recapitulated at the transcriptional level (Supplementary Table 2l–p). Comparisons of primary versus relapsed Group 4 recurrences (*n* = 3) show enrichment for gene sets involving extracellular matrix and cell surface receptor linked signal transduction. In contrast, Shh tumours (*n* = 3) show

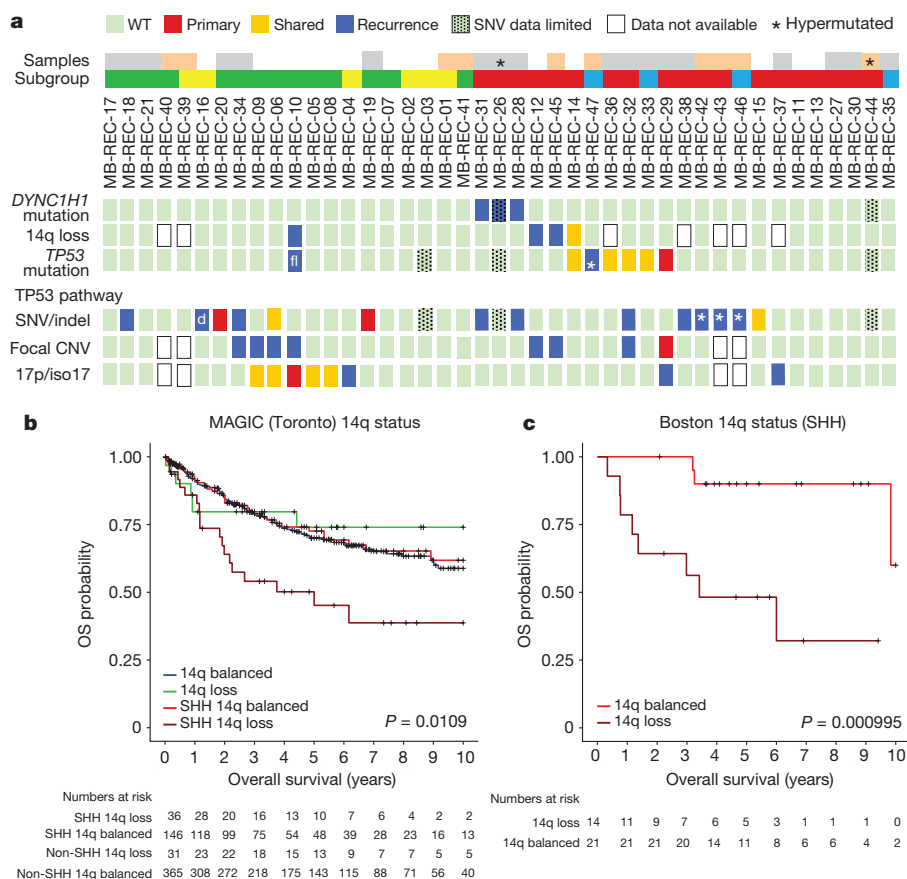


Figure 5 | Signalling pathways in recurrent medulloblastoma.

a, Compartment-specific deleterious events in the *TP53* gene ($n = 6/23$), genes from the *TP53* pathway ($n = 12/23$), *DYNC1H1* ($n = 3/23$), and chr14q loss (3/18). Asterisk indicates mutations in patients with missing diagnostic samples; 'd' indicates different events in pre- and post-therapy samples; white, patients with diagnostic, post-therapy, and germline

significant enrichment in P53 signalling and apoptosis-related gene sets, indicating that apoptotic escape may play a significant role in Shh recurrences; this is consistent with our therapy-resistant Shh mouse model, where we observed enrichment in *TP53* pathway gene sets in the local recurrences (Supplementary Table 1d).

Discussion

Genomic approaches, xenografts, and mouse models of medulloblastoma used experimentally to identify targets for rational therapy are based on untreated medulloblastoma. Paradoxically, novel agents are tested in children with highly treated medulloblastoma based on the assumption that biology at recurrence is largely similar to biology at presentation. Targeted therapy by definition is based on the identification of targets present exclusively in the diseased cell. Our data demonstrate that the model of tumour biology as static is not valid for medulloblastoma, which instead demonstrates striking evolution over time. Clinical trials of targeted therapy based on targets no longer present in the dominant clone at the time of recurrence would seem doomed to failure.

Although genetic events present in the dominant clone at diagnosis are unlikely to be present in the dominant clone at recurrence, we have previously shown that molecular subgroup affiliation is extremely stable at the time of recurrence, suggesting that therapeutic strategies based on susceptibility across a subgroup might be efficacious both upfront and at recurrence^{11,40}.

Our results comparing tumour genetics post-therapy for both human and murine medulloblastoma demonstrate that the dominant clone at recurrence arises at least in part through clonal selection of a minor clone that was already present at the time of diagnosis. Although our

samples; grey, no germline; pink, no matched diagnostic sample; blue, Wnt; red, Shh; yellow, Group 3; green, Group 4. **b**, Overall survival decreases in Shh patients with a chr14q-loss gene expression signature (versus balanced, log-rank test, $n = 578$, $P = 0.0109$); not significant in non-Shh tumours. **c**, Prognostic differences are replicated in an independent cohort (log-rank test, $n = 35$, $P = 0.000995$).

ability to identify recurrent driver SNVs and CNAs at the time of recurrence is hampered by sample size, there is a clear convergence in both human and mouse Shh medulloblastoma on events that drive genomic instability and aneuploidy. The recurrent convergence on a single pathway (genomic stability) after radiation treatment of Shh medulloblastoma suggests that it might be possible to develop anticipatory therapy⁴¹, in which genes/pathways responsible for therapeutic resistance are targeted at the time of initial therapy in order to prevent the emergence of resistance clones, or to modulate pathways such that resistant clones are outcompeted by therapy-sensitive ones⁴². Our 'humanized mouse model' of recurrent medulloblastoma demonstrates the remarkable power of appropriate 'humanized' model systems to predict pathways of therapy resistance. We would suggest that for all future clinical trials of targeted therapy of recurrent medulloblastoma, it should be mandatory to include re-biopsy to demonstrate maintenance of the target in the dominant recurrent clone. For cases in which the target is absent at recurrence, consideration should be given to using the novel agent in a neo-adjuvant manner.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 March; accepted 23 November 2015.

Published online 13 January 2016.

1. Hovestadt, V. *et al.* Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510**, 537–541 (2014).
2. Jones, D. T. *et al.* Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488**, 100–105 (2012).

3. Kool, M. *et al.* Genome sequencing of SHH medulloblastoma predicts genotype-related response to smoothened inhibition. *Cancer Cell* **25**, 393–405 (2014).
4. Northcott, P. A. *et al.* Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**, 428–434 (2014).
5. Northcott, P. A. *et al.* Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature* **488**, 49–56 (2012).
6. Pugh, T. J. *et al.* Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature* **488**, 106–110 (2012).
7. Wu, X. *et al.* Clonal selection drives genetic divergence of metastatic medulloblastoma. *Nature* **482**, 529–533 (2012).
8. Moxon-Emre, I. *et al.* Impact of craniospinal dose, boost volume, and neurologic complications on intellectual outcome in patients with medulloblastoma. *J. Clin. Oncol.* **32**, 1760–1768 (2014).
9. Northcott, P. A., Korshunov, A., Pfister, S. M. & Taylor, M. D. The clinical implications of medulloblastoma subgroups. *Nature Rev. Neurol.* **8**, 340–351 (2012).
10. Northcott, P. A. *et al.* Medulloblastomics: the end of the beginning. *Nature Rev. Cancer* **12**, 818–834 (2012).
11. Ramaswamy, V. *et al.* Recurrence patterns across medulloblastoma subgroups: an integrated clinical and molecular analysis. *Lancet Oncol.* **14**, 1200–1207 (2013).
12. Anderson, K. *et al.* Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature* **469**, 356–361 (2011).
13. Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**, 338–345 (2013).
14. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506–510 (2012).
15. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
16. Landau, D. A. *et al.* Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152**, 714–726 (2013).
17. Mullighan, C. G. *et al.* Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. *Science* **322**, 1377–1380 (2008).
18. Johnson, B. E. *et al.* Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science* **343**, 189–193 (2014).
19. Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809–813 (2009).
20. Gerlinger, M. & Swanton, C. How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *Br. J. Cancer* **103**, 1139–1143 (2010).
21. Eirew, P. *et al.* Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* **518**, 422–426 (2015).
22. Kreso, A. *et al.* Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science* **339**, 543–548 (2013).
23. Notta, F. *et al.* Evolution of human BCR-ABL1 lymphoblastic leukaemia-initiating cells. *Nature* **469**, 362–367 (2011).
24. Mumert, M. *et al.* Functional genomics identifies drivers of medulloblastoma dissemination. *Cancer Res.* **72**, 4944–4953 (2012).
25. Brett, B. T. *et al.* Novel molecular and computational methods improve the accuracy of insertion site analysis in Sleeping Beauty-induced tumors. *PLoS ONE* **6**, e24668 (2011).
26. Collier, L. S., Carlson, C. M., Ravimohan, S., Dupuy, A. J. & Largaespada, D. A. Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature* **436**, 272–276 (2005).
27. Dupuy, A. J., Akagi, K., Largaespada, D. A., Copeland, N. G. & Jenkins, N. A. Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature* **436**, 221–226 (2005).
28. Zhukova, N. *et al.* Subgroup-specific prognostic implications of TP53 mutation in medulloblastoma. *J. Clin. Oncol.* **31**, 2927–2935 (2013).
29. Zhu, S. *et al.* The bHLH repressor Deadpan regulates the self-renewal and specification of *Drosophila* larval neural stem cells independently of Notch. *PLoS ONE* **7**, e46724 (2012).
30. Ollmann, M. *et al.* *Drosophila* p53 is a structural and functional homolog of the tumor suppressor p53. *Cell* **101**, 91–101 (2000).
31. Fraley, C., Raftery, E. A., Murphy, T. B. & Scrucca, L. *mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation*. Technical Report No. 597 (2012).
32. Remke, M. *et al.* TERT promoter mutations are highly recurrent in SHH subgroup medulloblastoma. *Acta Neuropathol.* **126**, 917–929 (2013).
33. Govind, S. K. *et al.* ShatterProof: operational detection and quantification of chromothripsis. *BMC Bioinformatics* **15**, 78 (2014).
34. Griffith, M. *et al.* DGLdb: mining the druggable genome. *Nature Methods* **10**, 1209–1210 (2013).
35. McGranahan, N. & Swanton, C. Perspective biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell* **27**, 15–26 (2015).
36. Andor, N., Harness, J. V., Müller, S., Mewes, H. W. & Petritsch, C. EXPANDS: expanding ploidy and allele frequency on nested subpopulations. *Bioinformatics* **30**, 50–60 (2014).
37. Maley, C. C. *et al.* Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nature Genet.* **38**, 468–473 (2006).
38. Roth, A. *et al.* PyClone: statistical inference of clonal population structure in cancer. *Nature Methods* **11**, 396–398 (2014).
39. Hill, R. M. *et al.* Combined Myc and p53 defects emerge at medulloblastoma relapse and define rapidly progressive, therapeutically targetable disease. *Cancer Cell* **27**, 72–84 (2015).
40. Wang, X. *et al.* Medulloblastoma subgroups remain stable across primary and metastatic compartments. *Acta Neuropathol.* **129**, 449–457 (2015).
41. Aparicio, S. & Caldas, C. The implications of clonal genome evolution for cancer medicine. *N. Engl. J. Med.* **368**, 842–851 (2013).
42. Maley, C. C., Reid, B. J. & Forrest, S. Cancer prevention strategies that address the evolutionary dynamics of neoplastic cells: simulating benign cell boosters and selection for chemosensitivity. *Cancer Epidemiol. Biomarkers Prev.* **13**, 1375–1384 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements The MAGIC project is financially supported by: Genome Canada, Genome BC, Terry Fox Research Institute, Ontario Institute for Cancer Research, Pediatric Oncology Group Ontario, Funds from 'The Family of Kathleen Lorette' and the Clark H. Smith Brain Tumour Centre, Montreal Children's Hospital Foundation, Hospital for Sick Children: Sonia and Arthur Labatt Brain Tumour Research Centre, Chief of Research Fund, Cancer Genetics Program, Garron Family Cancer Centre, B.R.A.I.N. Child, and BC Childhood Cancer Parents Association. M.D.T. is also supported by a Stand Up To Cancer St. Baldrick's Pediatric Dream Team Translational Research Grant (SU2C-AACR-DT1113). Stand Up To Cancer is a program of the Entertainment Industry Foundation administered by the American Association for Cancer Research. M.D.T. is supported by The Canadian Cancer Society Research Institute, The Garron Family Chair in Childhood Cancer Research, and grants from the Cure Search for Children's Cancer Foundation, the National Institutes of Health (R01CA148699 R01CA159859), The Pediatric Brain Tumour Foundation, The Terry Fox Research Institute, Brainchild and The McLaughlin Centre at the University of Toronto. M.D.T. is also supported by the Swifty Foundation. L.G. was supported by the Davis M. Ferguson Memorial Fund at ABTA. Alex's Lemonade Stand Young Investigator Award supported V.R. This study was conducted with the support of the Ontario Institute for Cancer Research through funding provided by the Government of Ontario. This work was also supported by a Program Project Grant from the Terry Fox Research Institute, and a Grand Challenge Award from CureSearch for Children's Cancer. Additionally, this work was supported by the PedBrain Tumour Project contributing to the International Cancer Genome Consortium, funded by German Cancer Aid (109252) and by the German Federal Ministry of Education and Research (BMBF, grants 01KU1201A, MedSys 0315416C and NGFNplus 01GS0883). Funding by the German Childhood Cancer Foundation (Deutsche Kinderkrebsstiftung) to S.M.P., G.F. and T.P. The study was also financed by the Hungarian Brain Research Program Grant No. KTIA_13_NAP-A-V/3. and NAP-A-II/7. A.K. was supported by the János Bolyai scholarship of the Hungarian Academy of Sciences. E.G.V.M. was supported by NIH R01 grants CA163722 and NS096236, and St. Baldrick's and Cure Childhood Cancer Foundations. We would like to acknowledge R. P. Hill (Ontario Cancer Institute), the Labatt Brain Tumour Research Centre Tumour and Tissue Repository, which is supported by B.R.A.I.N. Child and Megan's Walk. M.R. is supported by a fellowship from the Dr. Mildred Scheel Foundation for Cancer Research/German Cancer Aid. F.M.G.C. is supported by the Stephen Buttrum Brain Tumour Research Fellowship, granted by Brain Tumour Foundation of Canada. V.R. is supported by a CIHR fellowship and an Alberta Innovates-Health Solutions Clinical Fellowship. We would like to thank the Toronto Centre for Phenogenomics for animal housing and veterinary support, and the Preclinical Core II and animal research facility at STARR (Spatiotemporal Targeting and Amplification of Radiation Response) in Toronto for assistance with CT-guided radiation experiments. We would like to thank Z. Wang for technical help with IHC, S. Archer for technical writing and C. Smith for artwork.

Author Contributions A.S.M., L.G., and M.D.T. led the study. L.G. planned and carried out *in vivo* and *in vitro* experiments and analyses, and performed a subset of bioinformatic analyses. A.S.M. supervised the RNA-seq and WGS experiments, led and executed bioinformatic analyses. D.J.H.S. performed bioinformatics analysis of mutation signatures. S.Z. developed and implemented the computational method of finding initiating events in mouse tumours. X.H. developed the *Drosophila* brain tumour model and performed imaging of *Drosophila* brains. P.S. assisted with mouse library preparation and bioinformatics analysis. M.R. and V.R. performed bioinformatics analyses on *DYNC1H1* and 14q loss. F.M.G.C. generated visualizations of structural rearrangements. P.E.L. and S.J. developed the radiotherapy schedule for the mouse model and designed the custom made collimators, beds, and stages for mouse CSI. K.Z. assisted with library preparation. B.L. extracted nucleic acids, managed the biobanking, and maintained the patient database. N.T., Y.M., and K.L.M. supervised bioinformatics analyses at the Genome Sciences Center, including sequence alignment, copy number analysis, and SNV and structural variant calling. Y.L., C.M. and E.M. performed bioinformatics analysis of human sequencing and deep-sequencing data. K.T. and T.Z. supervised and implemented the targeted deep-sequencing work. K.S. performed PyClone analysis. A.J.L.R. and S.S. designed and implemented PyClone, and supervised its use. H.F., S.M.-L., J.R., and T.P. assisted with bioinformatic analyses. J.L., and L.Q., assisted with animal care, and N.K., B.L.H., J.Y.L., L.K.D., Xin W., S.C.M., A.M., K.A.M., C.N., John P., A.R., and Y.Y.T. provided technical support. Xiaochong W. generated the transgenic mouse model and offered technical advice. A.A., M.B., Y.S.N.B., R.C., Y.C., E.C., R.C., N.D., A.H., D.L., H.I.L., W.L., M.M., P.P., J.Q.Q., J.E.S., A.T., T.W., I.B., and Y.Z., led and performed RNA-seq and WGS library preparation and sequencing experiments and performed data analyses. A.K., D.T.W.J., M.K., P.A.N., and S.M.P. at DKFZ performed the sequencing of four patients' sets.

C.C.F., José P., S.N., T.S., M.G., I.F.P., R.L.H., X.-N.Li, A.E.B., D.W.F., A.W.W., T.K., T.T., V.P.C., Y.-J.C., C.H., D.L., J.H.W., J.H.G. Jr, D.S.S., L.M., U.S., J.S., K.Z., S.P., O.A., S.E.D., D.P.C.T., C.G.C., H.W., A.R.H., W.I., T.J.M., J.J.O., E.G.V.M., J.-Y.L., K.-C.W., S.-K.K., B.-K.C., Y.S.R., S.B., J.C.L., S.C.C., C.G.E., M.K.C., R.J.P., M.M., M.L.G., N.J., and S.M.P. obtained the patient samples and clinical details that made the study possible. T.P., G.F., S.T., U.B., U.T., C.E.H., P.D., E.B., J.T.R., R.J.W.-R., W.A.W., L.S.C., A.J.D., A.K., D.T.W.J., M.K., P.A.N., S.M.P., D.A.L., A.J.M., R.A.M., N.J., G.D.B., S.J.M.J., and D.M. provided valuable input regarding study design, data analysis, and interpretation of results. A.S.M., L.G., M.R., S.Z., G.D.B., M.A.M. and M.D.T. wrote the manuscript. M.A.M. and M.D.T. provided financial and technical infrastructure and oversaw the study. M.A.M. and M.D.T. are joint senior authors and project co-leaders.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.D.T. (mdtaylor@sickkids.ca) or M.A.M. (mmarra@bcgsc.ca).

A. Sorana Morrissy^{1,2*}, Livia Garzia^{1,2*}, David J. H. Shih^{1,2,3}, Scott Zuyderduyn⁴, Xi Huang¹, Patryk Skowron^{1,2,3}, Marc Remke⁵, Florence M. G. Cavalli^{1,2}, Vijay Ramaswamy^{1,2,3,6}, Patricia E. Lindsay^{7,8}, Salomeh Jelveh⁸, Laura K. Donovan^{1,2}, Xin Wang^{1,2,3}, Betty Luu^{1,2}, Kory Zayne^{1,2}, Yisu Li⁹, Chelsea Mayoh⁹, Nina Thiessen⁹, Eloi Mercier⁹, Karen L. Mungall⁹, Yusanne Ma⁹, Kane Tse⁹, Thomas Zeng⁹, Karey Shumansky¹⁰, Andrew J. L. Roth¹⁰, Sohrab Shah¹⁰, Hamza Farooq^{1,2}, Noriyuki Kijima^{1,2}, Borja L. Holgado^{1,2}, John J. Y. Lee^{1,2,3}, Stuart Matan-Lithwick^{1,2}, Jessica Liu^{1,2}, Stephen C. Mack^{1,2,11}, Alex Manno^{1,2}, K. A. Michealraj^{1,2}, Carolina Nor^{1,2}, John Peacock^{1,2,3}, Lei Qin^{1,2}, Juri Reimand^{2,4}, Adi Rolider^{1,2}, Yuan Y. Thompson^{1,2,3}, Xiaochong Wu^{1,2}, Trevor Pugh^{1,2}, Adrian Ally⁹, Mikhail Bilenky⁹, Yaron S. N. Butterfield⁹, Rebecca Carlsen⁹, Young Cheng⁹, Eric Chuah⁹, Richard D. Corbett⁹, Noreen Dhalla⁹, An He⁹, Darlene Lee⁹, Haiyan I. Li⁹, William Long⁹, Michael Mayo⁹, Patrick Plettner⁹, Jenny Q. Qian⁹, Jacqueline E. Schein⁹, Angela Tam⁹, Tina Wong⁹, Inanc Biro^{9,13,14}, Yongjun Zhao⁹, Claudia C. Faria¹⁵, José Pimentel¹⁶, Sofia Nunes¹⁷, Tarek Shalaby¹⁸, Michael Grotzer¹⁸, Ian F. Pollack¹⁹, Ronald L. Hamilton²⁰, Xiao-Nan Li²¹, Anne E. Bendel²², Daniel W. Fu²³, Andrew W. Walter²⁴, Toshihiro Kumabe²⁵, Teiji Tominaga²⁶, V. Peter Collins²⁷, Yoon-Jae Cho²⁸, Caitlin Hoffman⁶, David Lyden²⁹, Jeffrey H. Wisoff³⁰, James H. Garvin Jr³¹, Duncan S. Stearns³², Luca Massimi³³, Ulrich Schüller³⁴, Jaroslav Sterba³⁵, Karel Zitterbart³⁵, Stephanie Puget³⁶, Olivier Ayrault³⁷, Sandra E. Dunn³⁸, Daniela P. C. Tirapelli³⁹, Carlos G. Carloti³⁹, Helen Wheeler⁴⁰, Andrew R. Hallahan^{41,42}, Wendy Ingram^{41,43}, Tobey J. MacDonald⁴⁴, Jeffrey J. Olson⁴⁵, Erwin G. Van Meir⁴⁶, Ji-Yeoun Lee⁴⁷, Kyu-Chang Wang⁴⁷, Seung-Ki Kim⁴⁷, Byung-Kyu Cho⁴⁷, Torsten Pietsch⁴⁸, Gudrun Fleischhack⁴⁹, Stephan Tippelt⁴⁹, Young Shin Ra⁵⁰, Simon Bailey⁵¹, Janet C. Lindsey⁵¹, Steven C. Clifford⁵¹, Charles G. Eberhart⁵², Michael K. Cooper⁵³, Roger J. Packer⁵⁴, Maura Massimino⁵⁵, Maria Luisa Garre⁵⁶, Ute Bartels⁵⁷, Uri Tabori^{2,57}, Cynthia E. Hawkins^{2,58}, Peter Dirks^{2,6}, Eric Bouffet^{2,57}, James T. Rutka^{2,3,6}, Robert J. Wechsler-Reya⁵⁹, William A. Weiss⁶⁰, Lara S. Collier⁶¹, Adam J. Dupuy⁶², Andrey Korshunov⁶³, David T. W. Jones⁶⁴, Marcel Kool⁶⁴, Paul A. Northcott⁶⁴, Stefan M. Pfister^{64,65}, David A. Largaespada⁶⁶, Andrew J. Mungall⁹, Richard A. Moore⁹, Nada Jabado⁶⁷, Gary D. Bader⁶⁸, Steven J. M. Jones^{9,13,69}, David Malkin^{57,70}, Marco A. Marra^{9,13§} & Michael D. Taylor^{1,2,3,6§}

¹Developmental & Stem Cell Biology Program, The Hospital for Sick Children, Toronto, Ontario M5G 0A4, Canada. ²The Arthur and Sonia Labatt Brain Tumour Research Centre, The Hospital for Sick Children, Toronto, Ontario, Canada. ³Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario M5G 0A4, Canada. ⁴The Donnelly Centre, University of Toronto, Toronto, Ontario M5S 3E1, Canada. ⁵Department of Pediatric Oncology, Hematology, and Clinical Immunology, University Hospital Düsseldorf, M5S 3E1, Germany. ⁶Division of Neurosurgery, The Hospital for Sick Children, Toronto, Ontario M5S 3E1, Canada. ⁷Department of Radiation Oncology, University of Toronto, Toronto, Ontario M5G 2M9, Canada. ⁸Radiation Medicine Program, Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario M5G 2M9, Canada. ⁹Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia V5Z 4S6, Canada. ¹⁰Department of Molecular Oncology, BC Cancer Agency, Vancouver, British Columbia V5Z 1L3, Canada. ¹¹Center for Stem Cell & Regenerative Medicine, Cleveland Clinic Foundation, Cleveland, Ohio 44195, USA. ¹²Clinical Genomics Research Program, Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario 44195, Canada. ¹³Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada. ¹⁴School of Computing Science, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada. ¹⁵Division of

Neurosurgery, Centro Hospitalar Lisboa Norte, Hospital de Santa Maria, Lisbon 1649-035, Portugal. ¹⁶Division of Pathology, Centro Hospitalar Lisboa Norte, Hospital de Santa Maria, Lisbon 1649-035, Portugal. ¹⁷Unidade de Neuro-Oncologia Pediátrica, Instituto Português de Oncologia de Lisboa Francisco Gentil, Lisbon 1099-023, Portugal. ¹⁸Departments of Oncology and Neuro-Oncology, University Children's Hospital of Zurich, Zurich 8032, Switzerland. ¹⁹Department of Neurological Surgery, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15224, USA. ²⁰Department of Pathology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15213, USA. ²¹Brain Tumor Program, Children's Cancer Center and Department of Pediatrics, Baylor College of Medicine, Houston, Texas 77030, USA. ²²Pediatric Hematology-Oncology, Children's Hospitals and Clinics of Minnesota, Minneapolis, Minnesota 55404, USA. ²³Department of Neurosurgery, Clinical Neurosciences Center, University of Utah, Salt Lake City, Utah 84132, USA. ²⁴A I duPont Hospital for Children, Wilmington, Delaware 19803, USA. ²⁵Department of Neurosurgery, Kitasato University School of Medicine, Sagami, Kanagawa 252-0374, Japan. ²⁶Department of Neurosurgery, Tohoku University Graduate School of Medicine, Sendai 980-8574, Japan. ²⁷Department of Pathology, University of Cambridge, Cambridge CB2 1QP, UK. ²⁸Departments of Neurosurgery, Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, California 94305, USA. ²⁹Departments of Pediatrics, Cell & Developmental Biology, Weill Medical College of Cornell University, New York, New York 10065, USA. ³⁰Department of Neurosurgery, NYU Langone Medical Center, New York, New York 10016, USA. ³¹Department of Pediatrics, Division of Pediatric Hematology, Oncology, and Stem Cell Transplantation, Columbia University, New York, New York 10032, USA. ³²Department of Pediatrics-Hematology and Oncology, Rainbow Babies & Children's Hospital and Department of Pediatrics-Hematology and Oncology, Case Western Reserve, Cleveland, Ohio 44106, USA. ³³Pediatric Neurosurgery, Catholic University Medical School, Rome 00198, Italy. ³⁴Center for Neuropathology, Ludwig-Maximilians-Universität, Munich 81377, Germany. ³⁵Department of Pediatric Oncology, School of Medicine, Masaryk University, Brno 602 00, Czech Republic. ³⁶AP-HP, Department of Neurosurgery, Necker-Enfants Malades Hospital, Université René Descartes, Paris 75743, France. ³⁷Signaling in Development and Brain Tumors, CNRS UMR 3347 / INSERM U1021, Institut Curie, Paris Cedex 5 91405, France. ³⁸Division of Hematology/Oncology, British Columbia Children's Hospital, Vancouver, British Columbia V6H 3V4, Canada. ³⁹Department of Surgery and Anatomy, Faculty of Medicine of Ribeirão Preto, Universidade de São Paulo, Brazil, Rebeirão Preto, São Paulo 14049-900, Brazil. ⁴⁰Kolling Institute of Medical Research, The University of Sydney, Sydney, New South Wales 2065, Australia. ⁴¹Queensland Children's Medical Research Institute, Children's Health Queensland, Brisbane, Queensland 4029, Australia. ⁴²Division of Oncology, Children's Health Queensland, Brisbane, Queensland 4029, Australia. ⁴³UQ Child Health Research Centre, The University of Queensland, Brisbane 4029, Australia. ⁴⁴Pediatric Neuro-Oncology Program, School of Medicine and Winship Cancer Institute, Emory University, Atlanta, Georgia 30307, USA. ⁴⁵Department of Neurosurgery, School of Medicine and Winship Cancer Institute, Emory University, Atlanta, Georgia 30322, USA. ⁴⁶Department of Hematology & Medical Oncology, School of Medicine and Winship Cancer Institute, Emory University, Atlanta, Georgia 30322, USA. ⁴⁷Department of Neurosurgery, Division of Pediatric Neurosurgery, Seoul National University Children's Hospital, Seoul 30322, South Korea. ⁴⁸Institute for Neuropathology, University of Bonn D-53105, Germany. ⁴⁹Children's University Hospital of Essen D-45147, Germany. ⁵⁰Department of Neurosurgery, University of Ulsan, Asan Medical Center, Seoul 05505, South Korea. ⁵¹Northern Institute for Cancer Research, Newcastle University, Newcastle upon Tyne NE1 4LP, UK. ⁵²Departments of Pathology, Ophthalmology and Oncology, John Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ⁵³Department of Neurology, Vanderbilt Medical Center, Nashville, Tennessee 37232-8550, USA. ⁵⁴Department of Neurology, Children's National Medical Center, Washington DC 20010-2970, USA. ⁵⁵Fondazione IRCCS Istituto Nazionale Tumori, Milan 20133, Italy. ⁵⁶U.O. Neurochirurgia, Istituto Giannina Gaslini, Genova 16147, Italy. ⁵⁷Department of Haematology & Oncology, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. ⁵⁸Division of Pathology, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. ⁵⁹Sanford-Burnham Medical Research Institute, La Jolla, California 92037, USA. ⁶⁰Departments of Pediatrics, Neurology and Neurosurgery, University of California San Francisco, San Francisco, California 94158, USA. ⁶¹School of Pharmacology, University of Wisconsin, Madison, Wisconsin 53715, USA. ⁶²Molecular & Cellular Biology Program, University of Iowa, Iowa City, Iowa 52242, USA. ⁶³Clinical Cooperation Unit Neuropathology, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany. ⁶⁴Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany. ⁶⁵Department of Pediatric Oncology, University Hospital Heidelberg, Heidelberg 69120, Germany. ⁶⁶Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota 55455, USA. ⁶⁷Division of Hematology/Oncology, McGill University, Montreal, Quebec H2W 1S6, Canada. ⁶⁸McLaughlin Centre and Department of Molecular Genetics, Banting and Best Department of Medical Research and Samuel Lunenfeld Research Institute at Mount Sinai Hospital, University of Toronto, Toronto, Ontario M5G 1L7, Canada. ⁶⁹Department of Molecular Biology & Biochemistry, Simon Fraser University, Burnaby, British Columbia M5G 1L7, Canada. ⁷⁰Department of Pediatrics, University of Toronto, Toronto, Ontario M5G 1X8, Canada.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

METHODS

No statistical methods were used to predetermine sample size.

Survival surgery for murine tumour removal. Male and female *Ptc^{+/-}/Math1-SB11/T2Onc* or *T2Onc2* mice (12 to 20 weeks of age; at the time they developed signs of medulloblastoma) were used. We did not perform a formal sample size estimate for the study but based our experimental plan on our previous experience with *Sleeping Beauty* mutagenesis screening. When mice showed early clinical signs of brain tumours they were anaesthetized with isoflurane, ophthalmic ointment applied to the eyes and the scalp antiseptically prepared. A 1.5 cm long midline incision was made to expose the skull from the coronal suture to the cranio-cervical junction. Using a high-speed drill and a 2.5 mm trephine bit, a cranial defect is drilled 2 mm posterior to lambda to avoid the transverse sinuses. The skull and the dura are lifted with micro-dissecting forceps, the bulk of the tumour is then removed using a harmon forceps with teeth, while smaller sections of tumour are removed with a microcurette (2 mm). Surgical samples are saved in dry ice, the bleeding from the tumour site is counteracted with direct pressure and Gelfoam. When haemostasis is obtained, the surgical wound is sutured using interrupted stitching with absorbable sutures. Animals received analgesia and dexamethasone post-operatively to contain the brain oedema. Male and female *Ptc^{+/-}/Math1-SB11/T2Onc* or *T2Onc2* control mice were monitored for early clinical signs of brain tumours but not subjected to surgery and CSI irradiation, no formal randomization was used. All the procedures involving animals have been approved by the institutional Animal Care Committee, in no case were tumour-bearing animals allowed a tumour burden compromising normal behaviour, food and water intake or exceeding the approved volume of 1,700 mm³.

CT guided craniospinal irradiation. Mice that had recovered from tumour resection were anaesthetized with isoflurane and placed in the brain irradiation bed in the image guided small animal irradiator (X-Rad 225CX, Precision Xray, North Branford, CT, USA). Correct animal setup was confirmed using 2D fluoroscopic images with and without the brain collimator (2 × 2 cm) in place, all images were acquired at 40 kVp, 0.5 mA, using the same X-ray tube which is used for radiation treatment. 3D volumetric cone-beam CT images were used for the visualization of bone and soft tissues within the animal and isocentre placement. The imaging capability of the unit were described previously⁴³, the imaging dose to the animal was estimated to be less than 1 cGy. The delivered dose per fraction was 2 Gy, administered 3 times a week for the first week to prevent brain oedema, followed by five times a week treatment for the following 3 weeks. Each daily dose was delivered with two parallel opposed-lateral beams to correct for tissue attenuation of the dose, total daily dose of 2 Gy. Dose rate for the brain collimator was measured at 3.2 Gy per min at 225 kVp, 13 mA, on a 0.3 mm Cu filter (HVL: 0.93 mm Cu, added filtration: 0.3 mm Cu). The tube was calibrated at these settings following the TG61 protocol⁴⁴.

The spine treatment was introduced on the second week of CSI irradiation, we used a 4.76 Gy per 6 fractions schedule, and the mice received 2 spinal fractions per week. Radiation to the spinal cord was delivered to mice placed supine on the irradiator stage the irradiation was done with single or multiple posterior beams. The same imaging strategy with 2D and volumetric 3D imaging was adopted for spinal cord targeting, using a 0.5 × 5 cm collimator or multiple fields of 0.5 × 2 cm; for the spine treatment a dose correction was applied to compensate for the different depth of the cervical spine compared to lumbo/sacral. Treatment dose was administered at 2.8 Gy per min at 225 kVp, 13 mA settings on a 0.3 mm Cu filter. The end-point date of the control and CSI treated groups was assessed by independent veterinary technicians blinded to the experimental group. Medulloblastoma-free survival from the time of diagnosis was assessed for control mice and mice that underwent surgery and radiation, no animal was removed from the study and mice euthanized during the study for different reasons than medulloblastoma were censored in the Kaplan–Meier estimate for tumour-free survival.

Linker-mediated PCR and Illumina HiSeq sequencing of transposon insertion sites. Genomic DNA was isolated and purified from mouse tissues with a PureLink genomic DNA extraction kit (Invitrogen). Libraries for Illumina HiSeq sequencing were prepared as described previously²⁵ with minor modifications. 2 µg of gDNA were digested and ligated to the adapters, after a BamHI digest to eliminate the untransposed copies of the concatamer, an enrichment PCR followed by a bar-coding PCR were performed²⁵. The barcode PCR was modified to incorporate a paired-end (PE) sequencing adaptor for paired end sequencing, the sequence of the PE adaptor was: CAAGCAGAAGACGGCATACGAGATCGGTCTCG GCATTCCTGCTGAACCGCTCTCCGATCTTAGGGCTCCGCTTAAGGGAC. Libraries were purified and pooled as previously described and sequenced on an Illumina HiSeq 2000 (ref. 25).

gCIS prediction. Sequenced libraries were demultiplexed and aligned as described previously²⁵. Demultiplexing and trimming of SB transposon sequences was performed using custom scripts, alignment of reads was performed with Novoalign

to mouse assembly NCBI37/mm9 (July 2007). A chi-squared test was used to assess statistical enrichment of the integration events within each transcription unit considering the following: the number of TA dinucleotide sites within the gene relative to the number of TA sites in the genome, the number of integration sites within each tumour, and the total number of tumours in each cohort. This gCIS analysis produced a *P* value for each of the 19,000 mouse RefSeq genes, and a Bonferroni correction was therefore used to adjust for multiple hypothesis testing. gCIS predictions were manually curated to filter out ambiguities, artefacts and local hopping. BioProject ID PRJNA306269.

Genetic algorithm for driver gene prediction. The model assumes that tumour cell division and growth are initiated by a founding transposon insertion event, and that additional insertion events can subsequently occur in daughter cells. According to the model, insertion events in the transformed daughter cells are expected to decrease by a factor of 2^{*n*} relative to the initial transformed cell, where *n* is the number of intervening cell divisions. Details of the model are described in Supplementary Note 1.

As with any model it is important to note its limitations. First, there is a limit to the degree to which distinct lineages can be separated. If two lineages are governed by two sufficiently close values of the parameter *G*, the components will be superimposed. If the value of *d* is also the same, the identification of the initiating insertions will not be affected; otherwise, the lineage with lower *d* will incorrectly identify its initiating mutation as a passenger. The extent of this issue is dependent upon the closeness of *G* and the depth at which a sample has been sequenced. It almost certainly true that other lineages are present in the data, but arose relatively late and/or have relatively low growth rates. Therefore, the model is best described as identifying the most clear and unambiguous lineages.

Second, a lineage which have undergone multiple gene disruptions that affect growth rate at different generations can appear as two separate lineages. For example, if a disruption of gene *A* causes rampant cell division/growth, and is followed up two generations later by a disruption in gene *B* that further increases the growth rate, this will appear as two lineages with putative genotypes *A*- and *B*-. In reality, the genotypes are *A*- and *A*;-*B*-. Importantly, this does not affect the ultimate identification of both of these genes as initiators.

Relative clonal prevalence was calculated for the genes predicted as driver as: 2^{*d*} *G* and normalized to the total number of predicted drivers for each sample. Driver events predicted to happen in the founder clones (highest *G*) for each sample, or showing relative cell abundance >10% were selected for pathway enrichment analysis.

PCR for *Sleeping Beauty* tagged fragments. The primers for amplifying *Sleeping Beauty* transposon insertion sites were designed based on the chromosomal location of each insertion site and its orientation to the transcription of the gene hosting the insertion. The primers at the inverted repeats/direct repeats (left) (IRL) and inverted repeats/direct repeats (right) (IRR) of the transposon were 5'-AAATTTGTGGAGTAGTTGAAAAACGA-3' and 5'-GGATTAAATGTCAGGAATTGTGAAA-3', respectively. The input represents genomic DNA with *Sleeping Beauty* transposition, which was illustrated by *Sleeping Beauty* excision PCR that detected the transposon post-transposition²⁶. Three points of input (1 ×, 5 × and 25 ×) were used.

Histology. Mice showing signs of late stage brain tumours were euthanized and tissue harvested for genomic DNA extraction as well as histological examination. Extent and location of recurrences was evaluated by standard haematoxylin and eosin staining. Trp53 pathway status was evaluated by p21 staining performed at the Paediatric Laboratory Medicine Department, The Hospital for Sick Children, (Toronto, Canada) using the Ventana BenchMark XT model. The conditions were as follows: HIER: 40 min in a Tris based buffer (pH 8.5) Ventana CC1 (<http://www.ventana.com/product/203?type=204>), primary antibody p21 (1:50) (BD bioscience 556431, clone SXM30) was incubated for 1 h at 37 °C. The signal was detected using Ventana OptiView DAB IHC Detection Kit.

Fly stocks. The following fly stocks were used: *UAS-mCD8-GFP* to label cell membrane; *insc-Gal4* (Gal41407 inserted in an *inscuteable* promoter) to drive gene expression in the neuroblast lineage; *UAS-dpn* for overexpression of *dpn*.

Fly culture. Flies were mated and maintained at 25 °C. Fly larvae were retrieved at late third instar stage for whole body irradiation at 40 Gy. The larval brains were dissected 4 h after irradiation, followed by fixation and immunohistochemistry analysis.

Immunofluorescence and imaging. Larval brains were dissected, fixed, and stained as previously described²⁹. Briefly, third instar larvae brains were dissected in PBS, fixed in 4% paraformaldehyde solution for 20 min at room temperature, and incubated with the primary antibody (rabbit anti-phospho-histone 3, Millipore, 1:200) overnight at 4 °C and secondary antibody for 2 h at room temperature. Fluorescence images were acquired using a Leica SP5 confocal microscope. Representative images of the dorsal brain lobes were shown in Fig. 1d, e and Extended Data Fig. 1h.

Whole-genome sequencing. All patients gave informed consent to the samples collection; unless indicated otherwise, the samples were sequenced and analysed at Canada's Michael Smith Genome Sciences Centre at the BC Cancer Agency (GSC). Libraries for whole-genome sequencing were constructed using either the plate-based or SPRI-TE library construction protocol.

Illumina genomic-plate-based library construction (350–450 bp insert size). 2 µg of genomic DNA in a 96-well format was fragmented by Covaris E210 sonication for 30 s using a 'duty cycle' of 20% and 'intensity' of 5. The paired-end sequencing library was prepared following the BC Cancer Agency's Genome Sciences Centre 96-well genomic ~350–450 bp insert Illumina Library Construction protocol on a Biomek FX robot (Beckman-Coulter, USA). Briefly, the DNA was purified in a 96-well microtitre plate using Ampure XP SPRI beads (40–45 µl beads per 60 µl DNA), and was subject to end-repair, and phosphorylation by T4 DNA polymerase, Klenow DNA Polymerase, and T4 polynucleotide kinase respectively in a single reaction, followed by cleanup using Ampure XP SPRI beads and 3' A-tailing by Klenow fragment (3' to 5' exo minus). After cleanup using Ampure XP SPRI beads, PicoGreen quantification was performed to determine the amount of Illumina PE adapters used in the next step of adaptor ligation reaction. The adaptor-ligated products were purified using Ampure XP SPRI beads, then PCR-amplified with Phusion DNA Polymerase (Thermo Fisher Scientific Inc. USA) using Illumina's PE indexed primer set, with cycle conditions: 98 °C for 30 s followed by 6 cycles of 98 °C for 15 s, 62 °C for 30 s and 72 °C for 30 s, and a final extension at 72 °C for 5 min. The PCR products were purified using Ampure XP SPRI beads, and checked with Caliper LabChip GX for DNA samples using the High Sensitivity Assay (PerkinElmer, USA). PCR products of the desired size range were gel purified (8% PAGE or 1.5% Metaphor agarose in an in-house custom built robot), and the DNA quality was assessed and quantified using an Agilent DNA 1000 series II assay and Quant-iT dsDNA HS Assay Kit using Qubit fluorometer (Invitrogen), then diluted to 8 nM. The final concentration was confirmed by Quant-iT dsDNA HS Assay before generating 100 bp paired-end reads on the Illumina HiSeq 2000/2500 platform using v3 chemistry.

SPRI-TE genomic library construction. Whole-genome libraries of patient samples medulloblastoma-REC-03, -04, -06, -11, -12, -18, -19, -22–24, -26–33 have been constructed using the SPRI-TE 300–600 bp fragment protocol as follows.

Genome libraries with fragment size ranges of approximately 400 bp were constructed on a SPRI-TE robot (Beckman Coulter, USA) according to the manufacturer's instructions (SPRIworks Fragment Library System I Kit, A84801). Briefly, 1 µg of genomic DNA in a 60 µl volume, and 96-well format, was fragmented by Covaris E210 sonication for 30 s using a 'duty cycle' of 20% and 'intensity' of 5. Up to 10 paired-end genome sequencing libraries were prepared in parallel using the SPRI-TE 300–600 bp size-selection program. Following completion of the SPRI-TE run the adaptor ligated library templates were quantified using a Qubit fluorometer. 5 ng of adaptor ligated template was PCR amplified using Phusion DNA Polymerase (Thermo Fisher Scientific, USA) and Illumina's PE indexed primer set, with cycle conditions: 98 °C for 30 s followed by 10 cycles of 98 °C for 15 s, 62 °C for 30 s and 72 °C for 30 s, and a final amplicon extension at 72 °C for 5 min. The PCR products were purified using Ampure XP SPRI beads, and checked with Caliper LabChip GX for DNA samples using the High Sensitivity Assay (PerkinElmer, USA). PCR products of the desired size range were purified using gel electrophoresis (8% PAGE or 1.5% Metaphor agarose gels in a custom built robot) and the DNA quality was assessed and quantified using an Agilent DNA 1000 series II assay and Quant-iT dsDNA HS Assay Kit using Qubit fluorometer (Invitrogen), then diluted to 8 nM. The final concentration was verified by Quant-iT dsDNA HS Assay before Illumina Sequencing before generating 100 bp paired-end reads on the Illumina HiSeq 2000/2500 platform using v2 or v3 chemistry.

WGS. Alignment. After marking chastity failed reads, paired-end 100 bp raw reads were aligned to the reference genome GRCh37-lite (http://www.bcgsc.ca/downloads/genomes/9606/hg19/1000genomes/bwa_ind/genome) with the Burrows–Wheeler Aligner (BWA; version 0.5.7)⁴⁵. Bam files were sorted with SAMTools (version 0.1.13) and merged using Picard MarkDuplicates.jar (version 1.71). The merged bam files were subsequently indexed with SAMTools index (version 0.1.17) and submitted to the European Genome-phenome Archive (EGAD00001000946).

German Cancer Research Centre (DKFZ). Patient samples medulloblastoma-REC-13-16 and medulloblastoma-REC-34-35 were processed at the DKFZ in Heidelberg as previously described².

Analysed DNA was isolated using using a Qiagen Allprep DNA/RNA/Protein Mini Kit. On average 125 mg of homogenized (TissueLyser, Qiagen) tumour tissue was used for isolation of analytes. The manufacturer's protocol was adapted to allow for DNA and total RNA (including miRNA) isolation. DNA from matching blood samples was extracted using Qiagen Blood and Cell Culture Midi Kit according to the manufacturer's protocol. After quality control of isolated DNA (gel electrophoresis), extracted nucleic acids were submitted for sequencing.

Paired-end (PE) DNA library preparation was carried out using Illumina Inc. v2 protocols. In brief, 1–5 µg of genomic DNA were fragmented to ~300 bp (PE) insert-size with a Covaris device, followed by size selection through agarose gel excision. Deep sequencing was carried out with Illumina HiSeq2000 instruments. **Whole-exome sequencing at McGill.** Patient samples medulloblastoma-REC-36-38 and medulloblastoma-REC-48-55 were prepared and sequenced by the Genome Quebec Innovation Centre and analysed at the McGill University Health Centre as follows. Paired-end libraries were prepared with the Illumina's Nextera Rapid Capture Exome kit. Captured exome DNA fragments were then sequenced on Illumina HiSeq 2500 (rapid-run mode) generating 100-bp paired-end reads. Adaptor sequences were removed and low-quality reads were trimmed using the FASTX toolkit. Quality trimmed reads were aligned to the human genome reference library (hg19) using Burrows–Wheeler Aligner (BWA) version 0.5.9 (ref. 45). Indels were realigned using the Genome Analysis Toolkit (GATK)⁴⁶ and duplicate reads were marked using Picard.

Genomic SNV analysis. SNVs from WGS data were analysed using all three methods described below, whereas SNVs from exome-seq data were analysed only with MutationSeq.

Samtools mpileup. SNVs were analysed with SAMtools mpileup v0.1.17 either on single or paired libraries. Each chromosome was analysed separately using the -C50-DSBuf parameters. The resulting vcf files were merged and filtered to remove low-quality SNVs by using samtools varFilter (with default parameters) as well as to remove SNVs with a QUAL score of less than 20. Finally, SNVs were annotated with gene annotations from Ensembl v66 using snpEff and the dbSNP v137 db membership assigned using SnpSift⁴⁷.

Strelka. To analyse compartment specific SNVs and indels, samples were analysed pair-wise with the default settings of Strelka v0.4.7 (ref. 48). Primary tumour samples and relapse/met were compared against the germline sample. In the absence of a germline sample, the relapse/met samples were compared against the primary tumour sample.

SNV classification. Variant allele frequencies (VAF) of somatic damaging SNVs (called by Strelka in 14 patients with matched germline samples) were classified into distinct clusters using the R package mclust, which uses finite mixture estimation via iterative expectation maximization steps (EM) and the Bayesian Information Criterion (BIC). Each cluster is manually categorized as either 'homozygous', 'clonal', or 'subclonal', depending on the cluster VAF and the uncertainty separating it from the next cluster. Multiple subclonal populations are numbered sequentially, starting with the most highly prevalent population.

MutationSeq. SNVs were analysed pair-wise with SAMtools mpileup v0.1.17 (ref. 49). Each chromosome was analysed separately using the -C50-DSBuf parameters. Before merging the resulting vcf files, they were filtered to remove all indels and low quality SNVs by using samtools varFilter (with default parameters) as well as to remove SNVs with a QUAL score of less than 20 (vcf column 6). The SNVs in the resulting vcf files were further filtered and scored using mutationSeq v1.0.2 and annotated with gene annotations from Ensembl v66 using SnpEff and the dbSNP v137 and Cosmic 64 db membership using SnpSift.

VarScan. Indels were called in the low quality exomes using VarScan version 2.3.6, using the following parameters: P value 95×10^{-2} –strand-filter 1 –min-avg-qual 20. The indels in the resulting vcf files were annotated with gene annotations from Ensembl 66 using SnpEff as described above, and screened against dbSNP137 using SnpSift.

Mutation spectra. EMu was used to define mutation spectra for 11 samples with germline (that is, excluding the DKFZ samples), using the expectation-maximization algorithm⁵⁰. To assess significant changes in the distributions of mutation spectra across primary, local and distal recurrences from each medulloblastoma patient, we used the chi-squared test. Changes in (1) the number of compartment-specific mutations and (2) in frequencies of transversion mutations, were tested with the Wilcoxon rank-sum test. Changes in the frequency of C > T and T > G transversions between primary and recurrent tumours were tested using factorial ANOVA with rank transformation.

CNV analysis (CNA-seq). The techniques outlined in ref. 51 were followed to analyse copy number changes. Sequence quality filtering was used to remove all reads of low mapping quality ($Q < 10$). Due to the varying amounts of sequence reads from each sample, aligned reference reads were first used to define genomic bins of equal reference coverage to which depths of alignments of sequence from each of the tumour samples were compared. This resulted in a measurement of the relative number of aligned reads from the tumours and reference in bins of variable length along the genome, where bin width is inversely proportional to the number of mapped reference reads. A hidden Markov model (HMM) was used to classify and segment continuous regions of copy number loss, neutrality, or gain using methodology outlined previously⁵². The five states reported by the HMM were: loss (1), neutral (2), gain (3), amplification (4), and high-level amplification

(5). In cases with germline, copy number gains and losses are called against the germline sample. In cases without germline, CNV calls were made using the primary instead of the germline sample, such that gains and losses reported in the recurrent tumour are relative to the copy number state in the primary. The limitations of this approach are that (1) when both primary and recurrent tumours share an event, the CNV output looks normal, and (2) when a gain (or loss) is called in the recurrent tumour versus the primary tumour, we cannot distinguish between the two scenarios that can give rise to such a result. The first scenario is that there is a gain the recurrence vs the primary, and the second is that there is a loss in the primary only. To resolve this uncertainty for particular chromosomes of interest in a subset of patients without germline, we additionally ran the Control-FREEC algorithm⁵³. Control-FREEC was run using the following default parameters, with the following exceptions: breakPointType = 4, telocentromeric = 75,000, minimalCoveragePerPosition = 5.

Structural variant detection. Structural variant detection was performed using ABySS (v1.3.2). Genome (WGS) libraries were assembled in single-end mode using *k*-mer values of k24, and k44. The contigs and reads were then reassembled at k64 in single end mode and then finally at k64 in paired end mode. Large-scale rearrangements and gene fusions were identified using BWA (v0.6.2-r126) alignments. Evidence for the alignments were provided from aligning reads back to the contigs and from aligning reads to genomic coordinates. Events were then filtered on read thresholds. Insertions and deletions were identified by gapped alignment of contigs to the human reference using BWA. Confidence in the event was calculated from the alignment of reads back to the event breakpoint in the contigs. The events were then screened against dbSNP and other variation databases to identify putative novel events.

SNV verification. To verify SNVs, samples were subjected to targeted deep amplicon sequencing of the tumour and normal DNA. Primers were designed with the Primer3 software⁵⁴ with a GC clamp and an optimal Tm of 64 °C to ensure specificity. Primers aligned against the human reference genome were tested with a combination of UCSC's *in silico* PCR tool and custom in-house scripts to obtain unique hits. The primer pairs were designed such that the variant is located within a maximum of 250 bp of the 5' or 3' amplicon end. The primers were tagged with Illumina adaptors eliminating the need for adaptor ligation during sample preparation. The Illumina adaptor tags are as follows: 5'-CGCTCTTCCGATCTCTG on the forward amplicon primer and 5'-TGCTCTTCCGATCTGAC on the reverse amplicon primers. Genomic DNA templates or library construction intermediates were used as starting material to generate PCR products using Phusion DNA polymerase (Fisher Scientific, catalogue number F-540L). The amplicons ranged in size from 188–625 bp. Amplicons were pooled by template for direct sequencing. Preparation for sequencing involved a second round of amplification (6 cycles with Phusion DNA polymerase) with PE primer 1.0-DS (5'-AATGATACGGCGACCGAGATCTACACTCTTCCCTA CACGACGCTCTTCCGATCTCTG-3') and a custom PCR primer (5'-CA AGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGT GTGCTCTTCCGATCTGAC-3') containing an unique six-nucleotide 'index' shown here as the letter N. PCR products of the desired size range were purified using 8% PAGE gels. DNA quality was assessed using the Agilent DNA 1000 series II assay (Agilent, Santa Clara CA, USA) and DNA quantity was measured using by Quant-iT dsDNA HS assay on a Qubit fluorometer (Life Technologies, Grand Island, NY, USA). The indexed libraries were pooled together and sequenced on the Illumina MiSeq platform with paired-end 250 bp reads using v2 reagents. An in-house generated PhiX sequencing control library was spiked in to the samples at molar ratio of 1:100. Reads were aligned using BWA-SW, and SNVs called with Samtools mpileup with the following parameters: -d 1000000 -B C50 -DES. Indels were called using VarScan and the following parameters: mpileup2indel-min-var-freq 0-p-value 1-strand-filter 0.

Detection of rare subclones in the primary tumours using DeepSeq data. SNVs with allelic frequencies greater than 15% in recurrent tumours were considered clonal. To find evidence for rare subclones (<5%) of these SNVs in the primary samples, we generated base quality (baseQ) distributions supporting the reference and all alternate alleles in the primary (and the recurrent) compartments. Due to our amplification and sequencing strategy, all reads start at the same position, and the target SNV is always at a specific position in the read (that is, a given mutation covered by 2,000 reads will be at base position 40 in all reads). Thus, unlike shotgun protocols where read starts are random, the SNVs are never affected by sequencing errors at the end of the read (where errors tend to happen more often), and cumulative sequencing error rates for whole reads are not applicable in estimating local error rates at a specific base. Instead, detection of a real mutation is only confounded by the subset of sequencing errors at the same position in the read that causes a base change to match the mutation; sequencing errors matching the other two possible bases (that is, non-reference and non-mutation) are a non-ambiguous

measure of the error rate at a particular position. Thus, to distinguish sequencing errors from real subclonal mutations, for each allele (that is, the reference allele and all three alternate alleles), we generated base quality (baseQ) distributions from all reads covering the position of the mutation; the reference base was further used as the benchmark distribution of a base without appreciable sequencing errors (Extended Data Fig. 8). The non-reference alleles that had the highest (1) mean baseQ value, (2) max baseQ value, and (3) highest number of reads with baseQ values >30, were considered real events. When all three criteria were not matched, the subclonal presence of the mutation could not be confirmed. At positions where these criteria were matched, the baseQ distributions of the alternate allele closely matched the baseQ distribution of the positive control reference base, could be easily distinguished from sequencing errors, and nearly always matched the expected mutation at that position, confirming the subclonal presence of the mutation in the diagnostic sample.

PyClone statistical inference of tumour cell populations using deep sequencing

SNV information. The allelic ratios are modelled using a binomial distribution and incorporated into the HMM Titan calculations, where the output is a list of copy number and LOH events. The Titan run for a tumour sample that has the lowest SDbw score is the optimal result and the corresponding number of clonal clusters is the optimal one—this copy number information was then chosen for use in further analysis. Minor and major copy number counts calculated from the optimal TitanCNA zygosity states were attached to the allele frequency information for each SNV and was used as input for Pyclone 0.12.3. PyClone was used to infer subclonal populations for all samples in each case. It introduces a framework that can analyse all samples from a single case in the same run improving accuracy of the inference. Pyclone outputs cellular frequencies and clonal cluster membership for each genomic position, accounting for confounding factors such as mutational genotype in the context of copy number changes. All Pyclone analyses were done using a multi-sample model and a beta-binomial distribution, with pre-calculated parental copy number inferred by TitanCNA.

Genome-wide phylogenetic analysis. Copy number and LOH information was called for 14 patients with matched germline samples using Control-FreeC⁵³, an algorithm that provides fractional copy number level for segments. Sensitive mutation calling was performed using muTect⁵⁵ and clonal and subclonal somatic mutations were shortlisted if there was adequate sequence coverage in both primary and relapse tumour compartments (10 reads minimum). Shortlisted mutations and copy number segments in areas of neutral heterozygosity were used as input to EXPANDS³⁶. Phylogenetic relationships between the subpopulations inferred by the EXPANDS algorithm in primary and recurrent tumours were generated using both SNV and copy number segments and the BIONJ algorithm. The inferred cellular prevalence values of each subpopulation was used to generate a Shannon Index value for each compartment³⁷.

Gene-expression signature. We identified 14q associated genes in Shh medulloblastoma using ANOVA in the Partek Genomics Suite. Gene expression profiles were analysed according to 14q status in samples from a previously published Toronto data set containing only SHH medulloblastoma samples ($n = 82$)³⁶ in a subset of cases with available SNP6 data^{5,57}. The top 20 ranking signature genes were applied using *k*-means clustering using the R2 platform (<http://hgserver1.amc.nl/cgi-bin/r2/main.cgi>) on a non-overlapping, independent gene expression profiling cohort from Boston⁵⁸ sub-selecting only SHH medulloblastomas. Survival differences were analysed using log-rank statistics and Kaplan–Meier estimates.

Messenger RNA library construction and sequencing. Two micrograms of total RNA samples were arrayed into a 96-well plate and polyadenylated (Poly(A)⁺) messenger RNA (mRNA) was purified using the 96-well MultiMACS mRNA isolation kit on the MultiMACS 96 separator (Miltenyi Biotec, Germany) with on-column DNaseI-treatment as per the manufacturer's instructions. The eluted poly(A)⁺ mRNA was ethanol precipitated and resuspended in 10 µl of DEPC-treated water with 1:20 SuperaseIN (Life Technologies, USA). First-strand cDNA was synthesized from the purified poly(A)⁺ mRNA using the Superscript cDNA Synthesis kit (Life Technologies, USA) and random hexamer primers at a concentration of 5 µM along with a final concentration of 1 µg µl⁻¹ actinomycin D, followed by Ampure XP SPRI beads on a Biomek FX robot (Beckman-Coulter, USA). The second strand cDNA was synthesized following the Superscript cDNA Synthesis protocol by replacing the dTTP with dUTP in dNTP mix, allowing the second strand to be digested using UNG (Uracil-N-Glycosylase, Life Technologies, USA) in the post-adaptor ligation reaction and thus achieving strand specificity. The cDNA was quantified in a 96-well format using PicoGreen (Life Technologies, USA) and VICTOR3V Spectrophotometer (PerkinElmer, Inc. USA). The quality was checked on a random sampling using the High Sensitivity DNA chip assay (Agilent). The cDNA was fragmented by Covaris E210 (Covaris, USA) sonication for 55 s, using a duty cycle of 20% and intensity of 5. Plate-based libraries were prepared following the BC Cancer Agency's Michael Smith Genome Sciences Centre

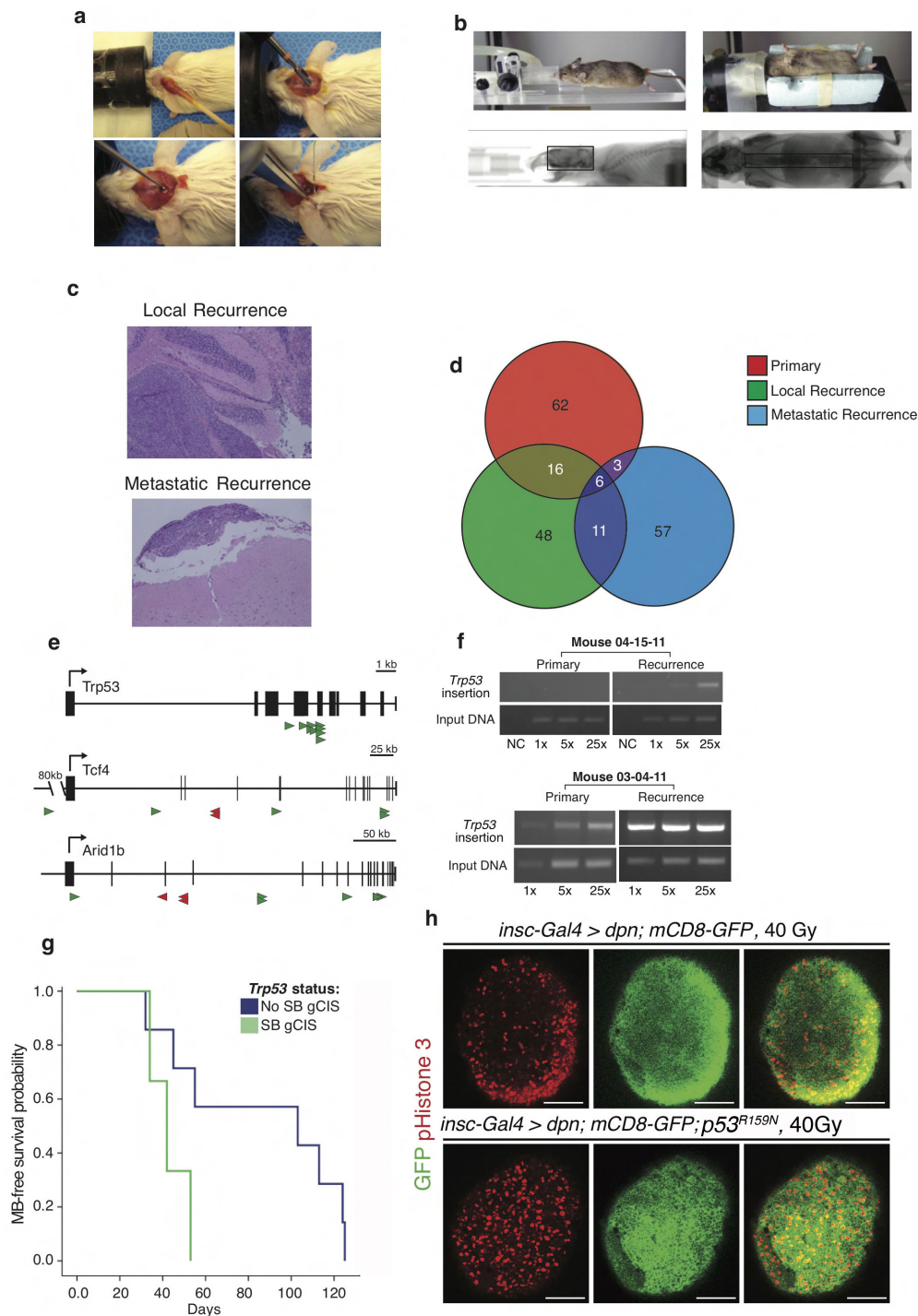
(BCGSC) paired-end (PE) protocol on a Biomek FX robot (Beckman-Coulter, USA). Briefly, the cDNA was purified in 96-well format using Ampure XP SPRI beads, and was subject to end-repair and phosphorylation by T4 DNA polymerase, Klenow DNA Polymerase, and T4 polynucleotide kinase respectively in a single reaction, followed by cleanup using Ampure XP SPRI beads and 3' A-tailing by Klenow fragment (3' to 5' exo minus). After cleanup using Ampure XP SPRI beads, PicoGreen quantification was performed to determine the amount of Illumina PE adapters used in the next step of adaptor ligation reaction. The adaptor-ligated products were purified using Ampure XP SPRI beads, then PCR-amplified with Phusion DNA Polymerase (Thermo Fisher Scientific USA) using Illumina's PE primer set, with cycle conditions of 98 °C 30 s followed by 10–15 cycles of 98 °C for 10 s, 65 °C for 30 s and 72 °C for 30 s, and then 72 °C for 5 min. The PCR products were purified using Ampure XP SPRI beads, and checked with a Caliper LabChip GX for DNA samples using the High Sensitivity assay (PerkinElmer, USA). PCR products with a desired size range were purified using a 96-channel size selection robot developed at the BCGSC, and the DNA quality was assessed and quantified using an Agilent DNA 1000 series II assay and Quant-iT dsDNA HS Assay Kit using Qubit fluorometer (Invitrogen), then diluted to 8 nM. The final concentration was verified by Quant-iT dsDNA HS assay. The libraries, 2 × 100 PE lanes, were sequenced on the Illumina HiSeq 2000/2500 platform using v3 chemistry and HiSeq Control Software version 2.0.10.

Alignment of strand-specific RNA-seq data. Illumina paired-end RNA sequencing data was aligned to GRCh37-lite genome-plus-junctions using BWA (version 0.5.7)^{49,59}. This reference is a combination of GRCh37-lite assembly and exon-exon junction sequences with coordinates defined based on transcripts in Ensembl (v61), Refseq and known genes from the UCSC genome browser (both were downloaded from UCSC in November 2011; The GRCh37-lite assembly is available at http://www.bcgsc.ca/downloads/genomes/9606/hg19/1000genomes/bwa_ind/genome). BWA "aln" and "sampe" were run with default parameters, except for the inclusion of the (-s) option to disable the Smith-Waterman alignment, which is unsuitable for insert size distribution in paired-end RNA-seq data. Finally, reads failing the Illumina chastity filter are flagged with a custom script, and duplicated reads were flagged with Picard Tools (version 1.31). After the alignment, the junction-aligned reads that mapped to exon-exon junctions were repositioned to the genome as large-gapped alignments and tagged with "ZJ:Z"⁵⁹.

Differential expression analysis of RNA-seq data. We compared the expression values (RPKM) of genes in the primary and recurrent tissues of each tumour with data in both compartments ($n = 7$ patients). A gene was considered differentially expression when the absolute difference between compartments was greater than 10 and the log₂ fold-change was greater than 2. Gene sets enrichment analysis was

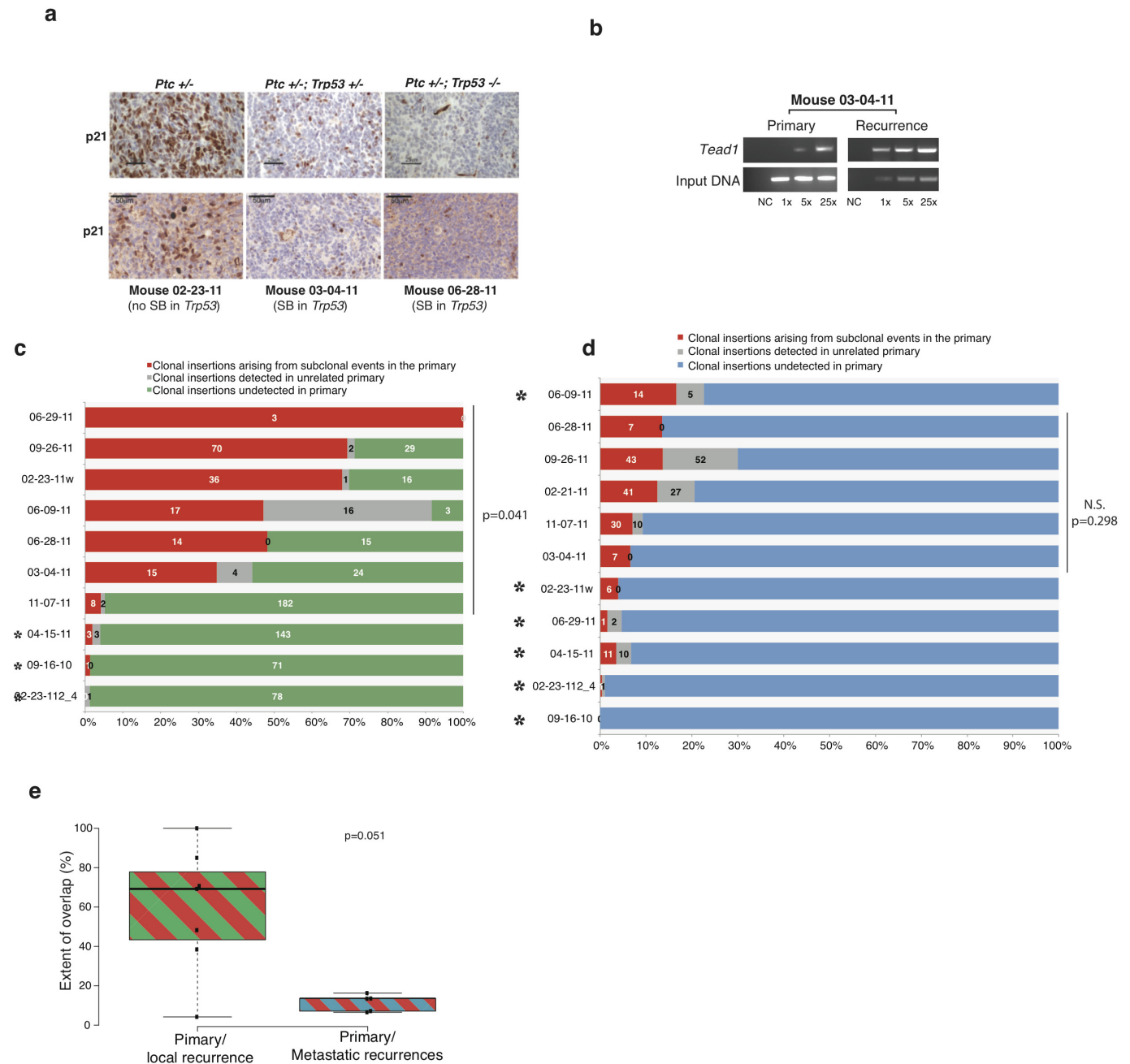
run on differentially expressed genes that were observed in at least two patients by subgroup, using mSigDB⁶⁰.

43. Clarkson, R. *et al.* Characterization of image quality and image-guidance performance of a preclinical microirradiator. *Med. Phys.* **38**, 845–856 (2011).
44. Ma, C. M. *et al.* AAPM protocol for 40–300 kV X-ray beam dosimetry in radiotherapy and radiobiology. *Med. Phys.* **28**, 868–893 (2001).
45. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
46. McKenna, A. *et al.* The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
47. Cingolani, P. *et al.* Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Front. Genet.* **3**, <http://dx.doi.org/10.3389/fgene.2012.00035> (2012).
48. Saunders, C. T. *et al.* Strelka: Accurate somatic small-variant calling from sequenced tumour-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
49. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
50. Fischer, A., Illingworth, C. J., Campbell, P. J. & Mustonen, V. EMU: probabilistic inference of mutational processes and their localization in the cancer genome. *Genome Biol.* **14**, R39 (2013).
51. Jones, S. J. *et al.* Evolution of an adenocarcinoma in response to selection by targeted kinase inhibitors. *Genome Biol.* **11**, R82 (2010).
52. Shah, S. P. *et al.* Integrating copy number polymorphisms into array CGH analysis using a robust HMM. *Bioinformatics* **22**, 431–439 (2006).
53. Boeva, V. *et al.* Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**, 423–425 (2012).
54. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386 (2000).
55. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnol.* **31**, 213–219 (2013).
56. Vanner, R. J. *et al.* Quiescent Sox2⁺ cells drive hierarchical growth and relapse in sonic hedgehog subgroup medulloblastoma. *Cancer Cell* **26**, 33–47 (2014).
57. Shih, D. J. H. *et al.* Cytogenetic prognostication within medulloblastoma subgroups. *J. Clin. Oncol.* **32**, 886–896 (2014).
58. Cho, Y.-J. *et al.* Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. *J. Clin. Oncol.* **29**, 1424–1430 (2011).
59. Butterfield, Y. S. *et al.* JAGuar: junction alignments to genome for RNA-seq reads. *PLoS ONE* **9**, e102398 (2014).
60. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).



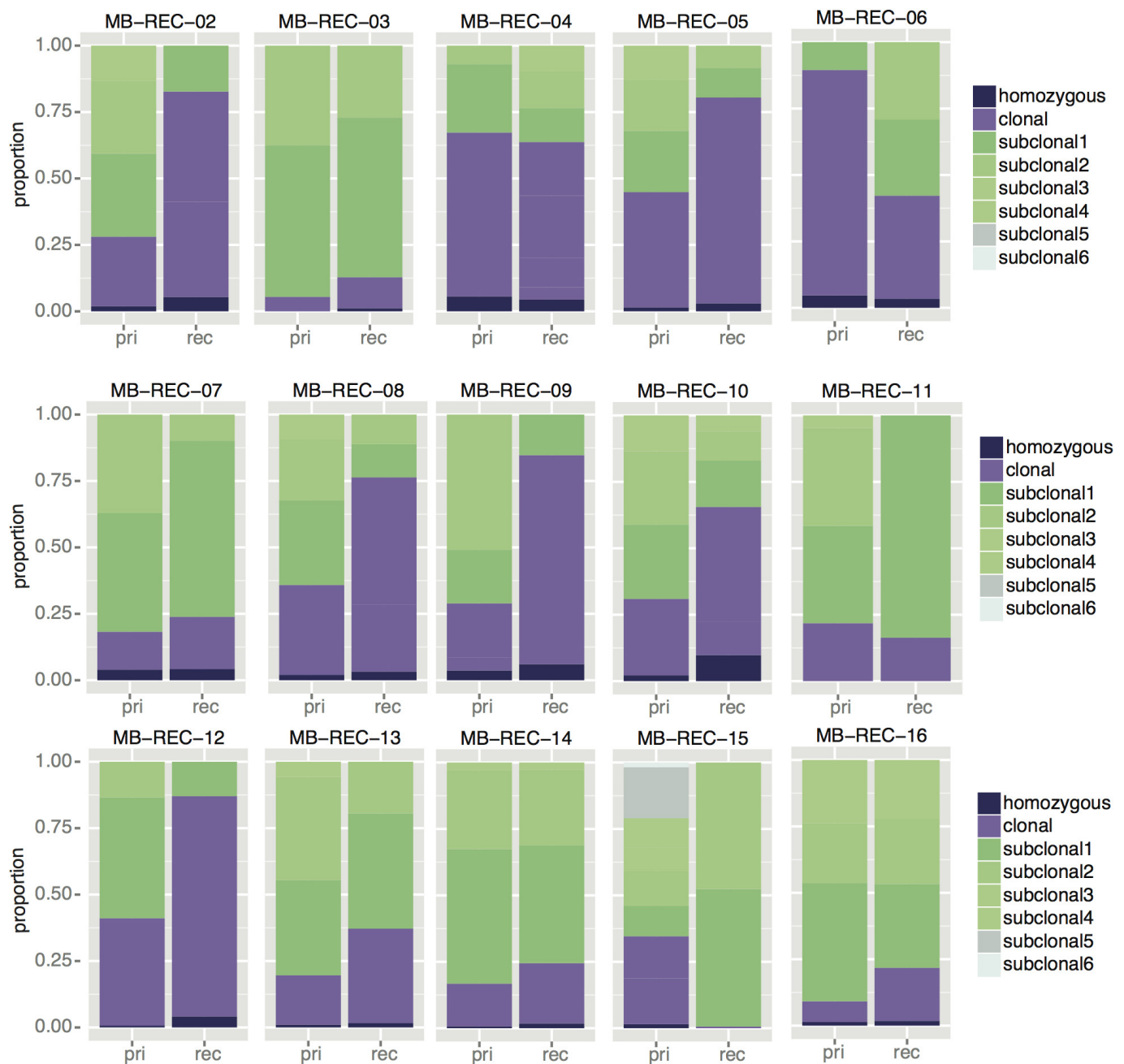
Extended Data Figure 1 | Microneurosurgical resection and CT guided multi-fractionated craniospinal radiotherapy in a Shh mouse model of medulloblastoma. **a**, Under general anaesthesia, *Ptch*^{+/-}/*Math1-SB11/T2Onc* mice with symptomatic medulloblastoma underwent microneurosurgical posterior fossa craniotomy and subtotal tumour removal ($n = 38$), followed by post-operative care and monitoring. **b**, Subsequently, post-operative mice are recurrently anaesthetized, and receive multi-fractionated cranial and spinal cord irradiation in 18 fractions for a total of 36 Gy over a period of four weeks. Radiation is delivered under computed tomography (CT) guidance using custom-made mouse beds and collimators in order to precisely target the entire craniospinal axis. **c**, Mice that completed the entire course of craniospinal radiation were cured of disease in 39% of cases (7/18), while the remainder had to be euthanized as they recurred locally and/or with leptomeningeal metastases (61%, 11/18). Histology (haematoxylin and eosin staining) at the time of autopsy is shown. **d**, Extent of overlap of primary, local recurrences and metastatic recurrences initiator genes as predicted by a

per-mouse driver modelling approach. **e**, Clonal transposon insertions in *Trp53*, *Tcf4* and *Arid1b* disrupt the coding sequence of the gene. Sense orientation insertions are illustrated in green, antisense insertions in red. **f**, Insertion-site end-point PCR demonstrates *Trp53* insertions that are clonal in the recurrence, but present only in a subclone of the matched primary tumour or completely absent. Three levels of input DNA were used for each sample (1x, 5x and 25x). **g**, Mice treated with microneurosurgical resection and craniospinal radiation, whose tumours show *Trp53* gCIS insertions in the local recurrence show a trend for a shorter survival than similarly treated mice without *Trp53* insertions (log-rank test; $P = 0.054$; $n = 10$). **h**, *Drosophila* brain tumours are induced by expressing *dpn* in the neural stem cell lineage using *insc-Gal4*. In response to a systemic 40 Gy irradiation at late third instar stage, overexpressing a dominant negative form of *Drosophila* p53, *p53*^{R159N}, resulted in moderately increased mitosis in tumour cells labelled by the membrane GFP (mCD8-GFP), scale bar, 50 μ m.



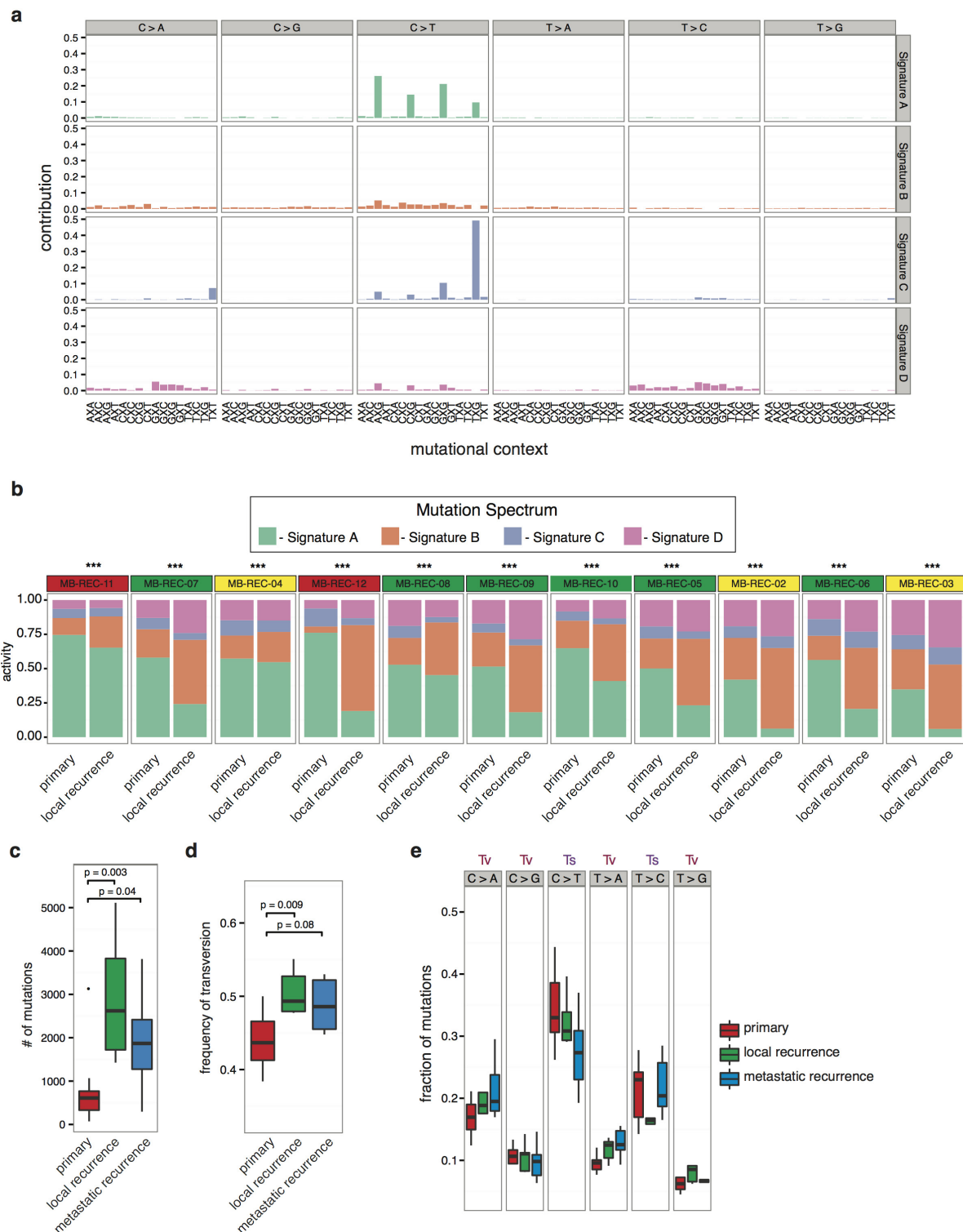
Extended Data Figure 2 | Subclonal events in primary mouse tumours become clonal at recurrence. **a**, Naive tumours from *Ptch*^{+/-}, *Ptch*^{+/-}/*Trp53*^{+/-} or *Ptch*^{+/-}/*Trp53*^{-/-} germline mutant mice were analysed by immunohistochemical staining for nuclear p21 (upper panels), demonstrating decreased nuclear p21 expression due to *Trp53* pathway dysfunction. Tumours with *Trp53* damaging gCIS insertions at recurrence (03-04-11 and 06-28-11) also show decreased immunohistochemical staining for nuclear p21 staining (lower panels), when compared to a recurrent tumour without gCIS *Trp53* insertions (02-23-11w) (scale bars, 25 μm and 50 μm as indicated). **b**, Relative dominance of driver events is shown in one individual tumour where *Tead1* is detectable in both primary tumour sample and at recurrence. **c**, **d**, Clonal insertions in the local and metastatic recurrences that were found at a subclonal level in the matching primary tumours are shown by mouse. In each case, the number of insertions with evidence for expansion from a subclone of the primary is shown as a proportion (red bar) of the total number of considered events. Green and blue bars depict the proportion of the total number of considered events that were found in local and metastatic recurrences, respectively. The grey bar indicates the proportion of insertions that are

also found in an unrelated *Sleeping Beauty* library of similar depth. We narrowed the analysis on matching primary recurrences with at least 1 clonal insertion in common. This excluded 3 local recurrence cases and 6 metastatic recurrence cases that had no overlap between clonal insertions in the primary and clonal insertions in the matching local recurrences (black stars). **c**, Local recurrences display statistical support for subclonal derivation from the primary tumours ($P = 0.041$; $n = 7$; Mann–Whitney *U*-test). **d**, Metastatic recurrences instead show a limited extent of overlap with the matching primaries that does not reach statistical significance ($P = 0.298$; $n = 5$; Mann–Whitney *U*-test). **e**, Box plot comparing the extent of overlap between primary/local recurrences versus primary/metastatic recurrences, local recurrences (with at least 1 clonal insertion in common with the primary) show a trend for higher evidence of subclonal derivation from their matched primaries than metastatic recurrences ($P = 0.051$, Mann–Whitney *U*-test, centre lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, samples are represented by dots. $n = 7$ and 5 sample points).



Extended Data Figure 3 | Subclonal events in primary human tumours become clonal at recurrence. The proportion of somatic SNVs in the primary and recurrent disease compartments of 15 patients with matched germline is shown as a function of clonality. Black indicates homozygous events, purple indicates clonal SNVs, and subclonal SNVs are shown in

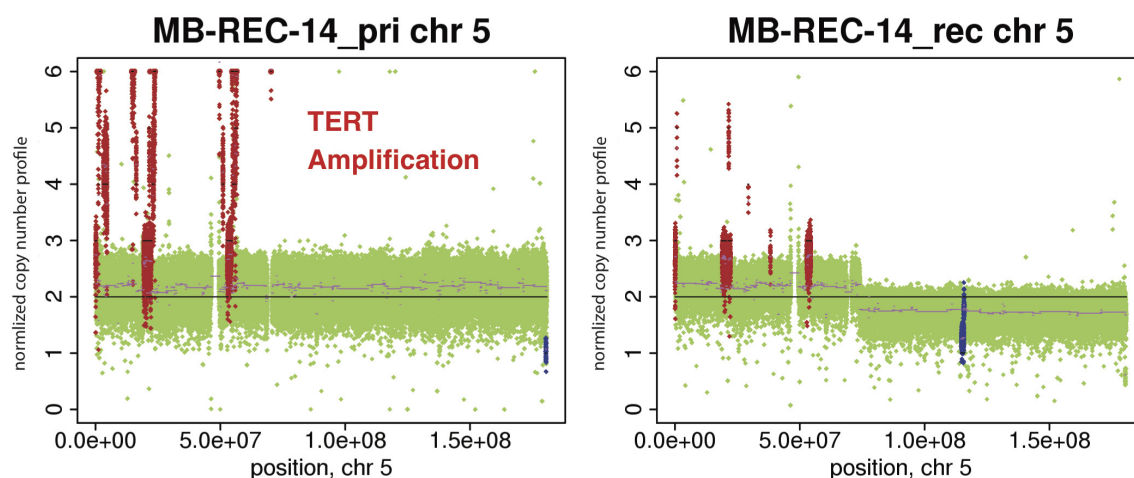
green, where lighter shades correspond to less abundant subpopulations. On average, we observe a 1.9-fold increase in the proportion of clonal and homozygous events across the cohort (Student's *t*-test; P value = 8.7×10^{-3} , $n = 15$).



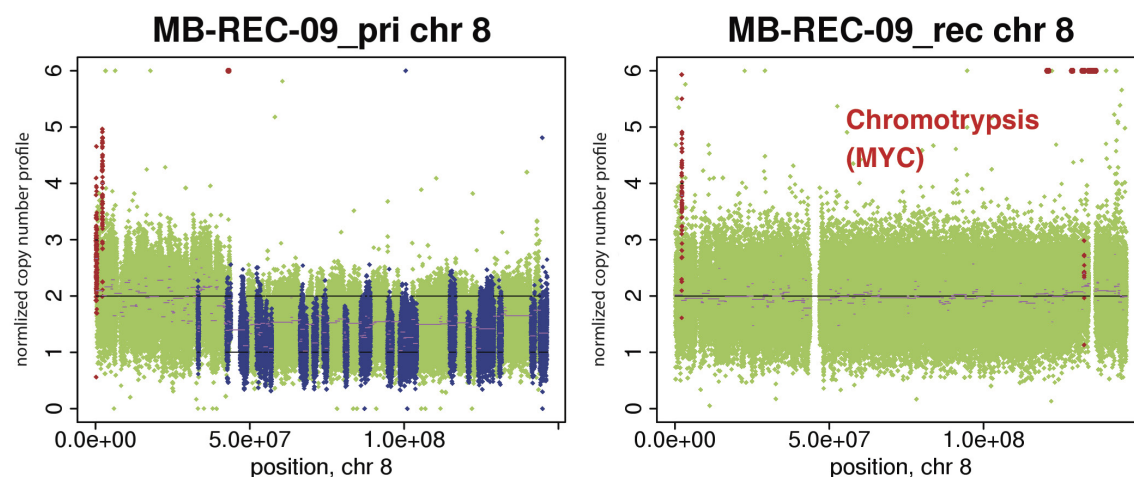
Extended Data Figure 4 | Altered spectra of somatic SNVs when comparing therapy-naïve to recurrent tumours. **a**, Mutations in each tumour sample ($n = 15$) were classified based on their sequence context, and clustered into signatures that represent four known mutational processes. Signature A is the age-related signature observed in most tumour types (deamination of methyl-C). Signature B is characterized by C > A and C > T mutations without a strict context requirement. Signatures C and D respectively resemble the MSI-L and MSI-H signatures that correlate with low (MSI-L) or high (MSI-H) microsatellite instability. **b**, The contribution of each mutational process to each primary and recurrent tumour is summarized by patient. Recurrent tumours show a shift away from signature A, and an increased prevalence of signature B and signature D. *** $P < 0.001$, chi-squared test denotes significantly

different distributions; NS denotes not significant. All tumours shifted mutational signatures at the time of recurrence, for **a** and **b**, $n = 15$. **c**, **d**, The number **c**, and frequency **d**, of transversion mutations is summarized in therapy-naïve and recurrent samples. Significant increases in the number and frequency of transversions is most strongly observed in local recurrences, and to a lesser extent in metastatic recurrences. $P < 0.05$, Wilcoxon rank-sum test. **e**, Breakdown of transversion (Tv) and transition (Ts) mutations in therapy-naïve and recurrent samples does not show a significant trend in specific nucleotide changes. Centre lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend up to 1.5 times the interquartile range (from the 25th to the 75th percentiles), and data points beyond the whiskers are outliers represented by dots. For **c**, **d** and **e**, $n = 13$, $n = 7$ and $n = 6$, respectively.

a



b



c

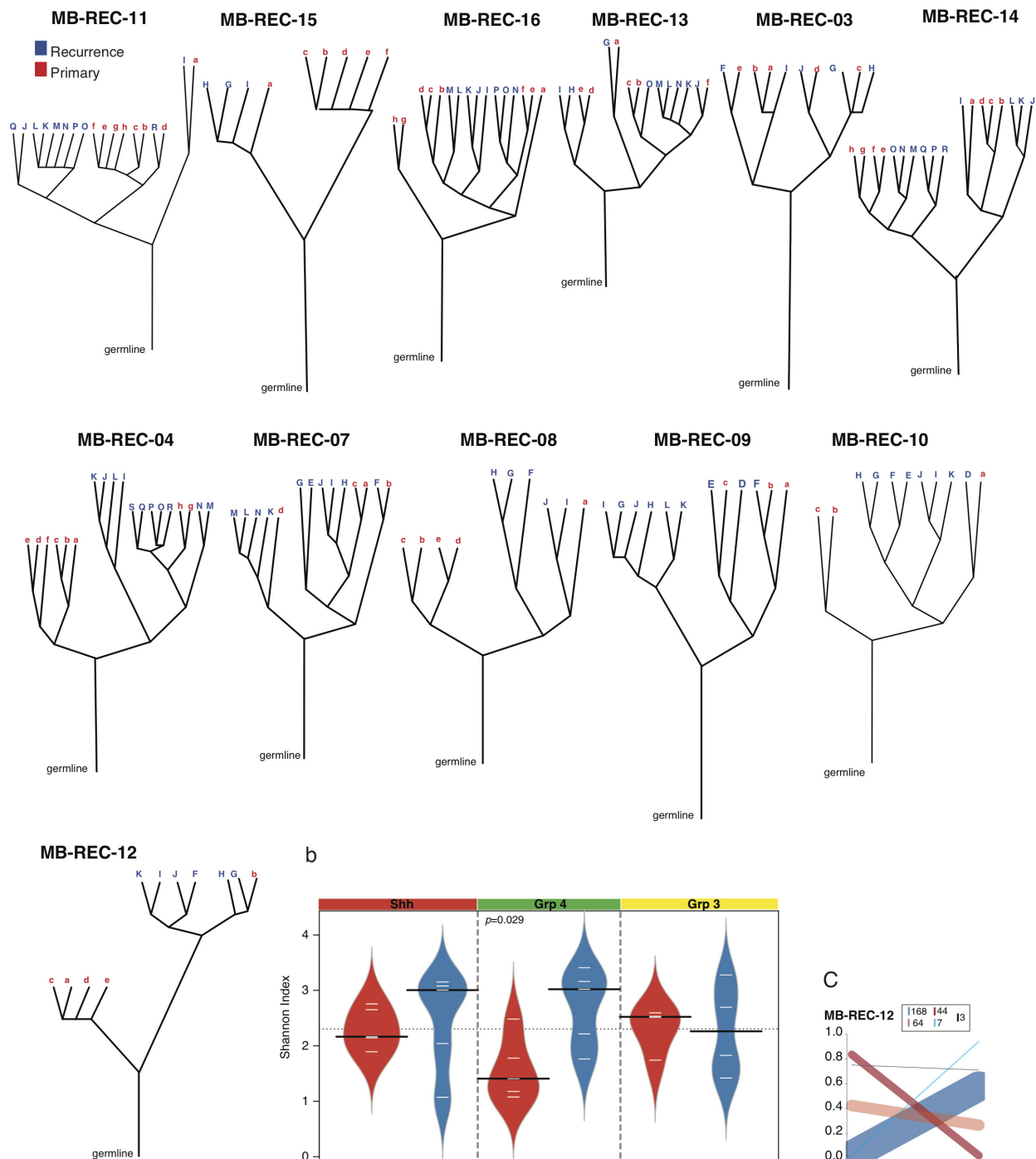
Druggable Targets (DGIdb neoplastic drugs only)

Patient_ID	Subgroup	Naïve Tumor	Shared	Post-Therapy Tumor
MB-REC-02	Group3	-	<i>RAF1</i>	EGFR
MB-REC-03	Group3	-	-	-
MB-REC-04	Group3	-	-	-
MB-REC-05	Group4	-	-	-
MB-REC-06	Group4	-	<i>DEPDC6,PTK2,TP53,ZHX2</i>	ANGPT1
MB-REC-07	Group4	-	-	FGFR4
MB-REC-08	Group4	-	-	<u>TP53</u>
MB-REC-09	Group4	-	-	<u>TP53, YES1</u>
MB-REC-10	Group4	-	-	FLT1
MB-REC-11	SHH	-	SMO*	DDR2
MB-REC-12	SHH	NOS2,PTCH1	-	<u>PDGFRA,PSME3,TP53</u>
MB-REC-13	SHH	-	<i>CRTC1,DDR2</i>	-
MB-REC-14	SHH	-	<i>MAP2K6,TP53</i>	-
MB-REC-15	SHH	-	<i>MAP2K7,PTCH1,PTEN,TP53</i>	HDAC4
MB-REC-16	Group3	ITGB1,PIK3C2G,TOP2B	-	HDAC7,TP53

Extended Data Figure 5 | Compartment-specific driver and druggable events in human tumours. a, High-level *TERT* amplification in the primary tumour of patient MB-REC-14 is absent in the recurrent sample. b, Chromothripsis involving the *MYC* locus is specific to the recurrent tumour on patient MB-REC-09 (P value = 3.97×10^{-7}). c, Genes with defined interactions to neoplastic drugs (DGIdb <http://dgidb.genome.wustl.edu/>). The majority of patients ($n = 15$; with matched or parental

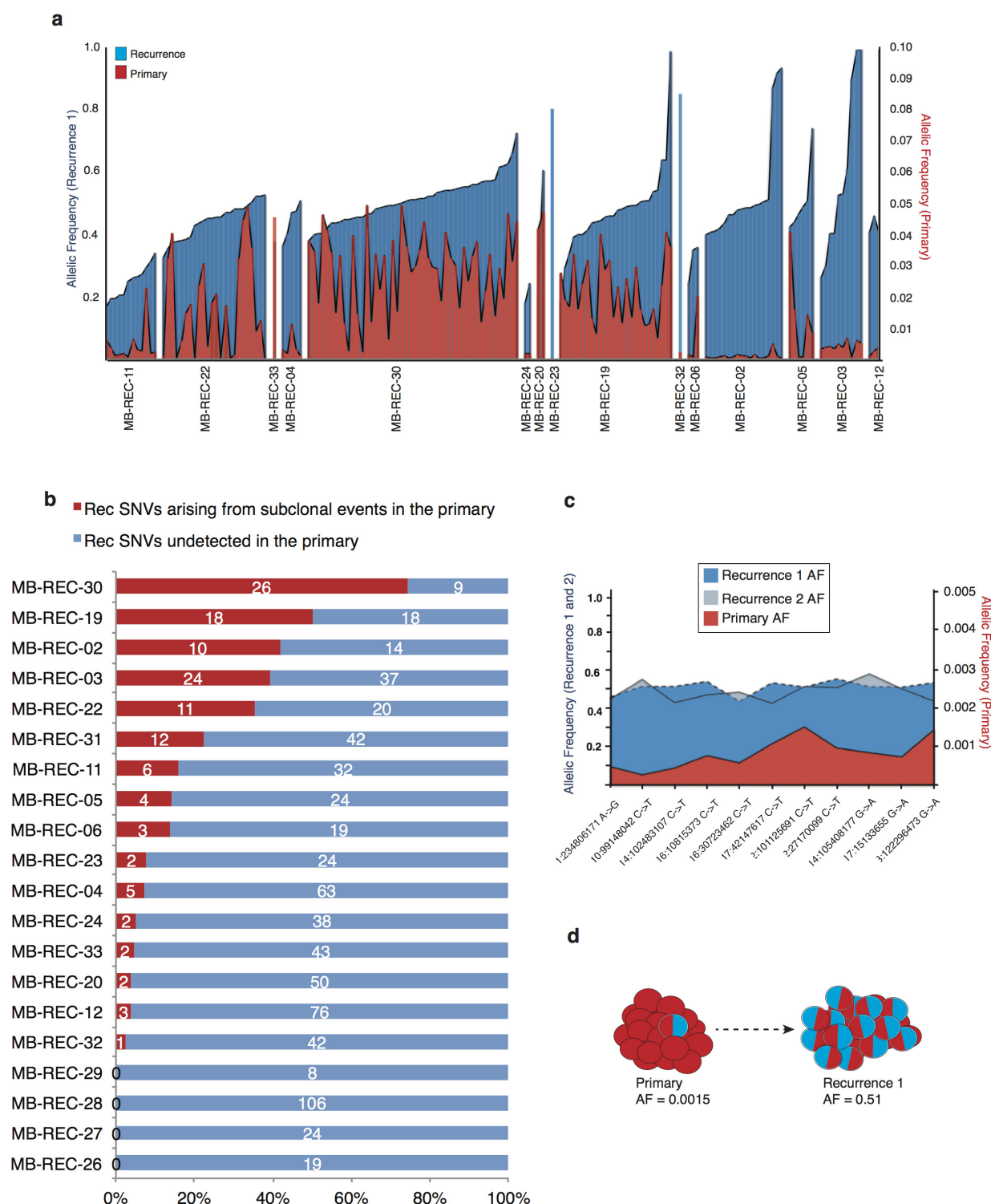
germline) have distinct druggable targets in the naïve versus post-therapy tumour samples. Bolded gene names indicate the presence of damaging mutations that are clonal (versus subclonal events in lighter colours), underlined gene names indicate copy number aberrations (for example, loss at the *TP53* locus), and italicized gene names indicate structural rearrangements.

a



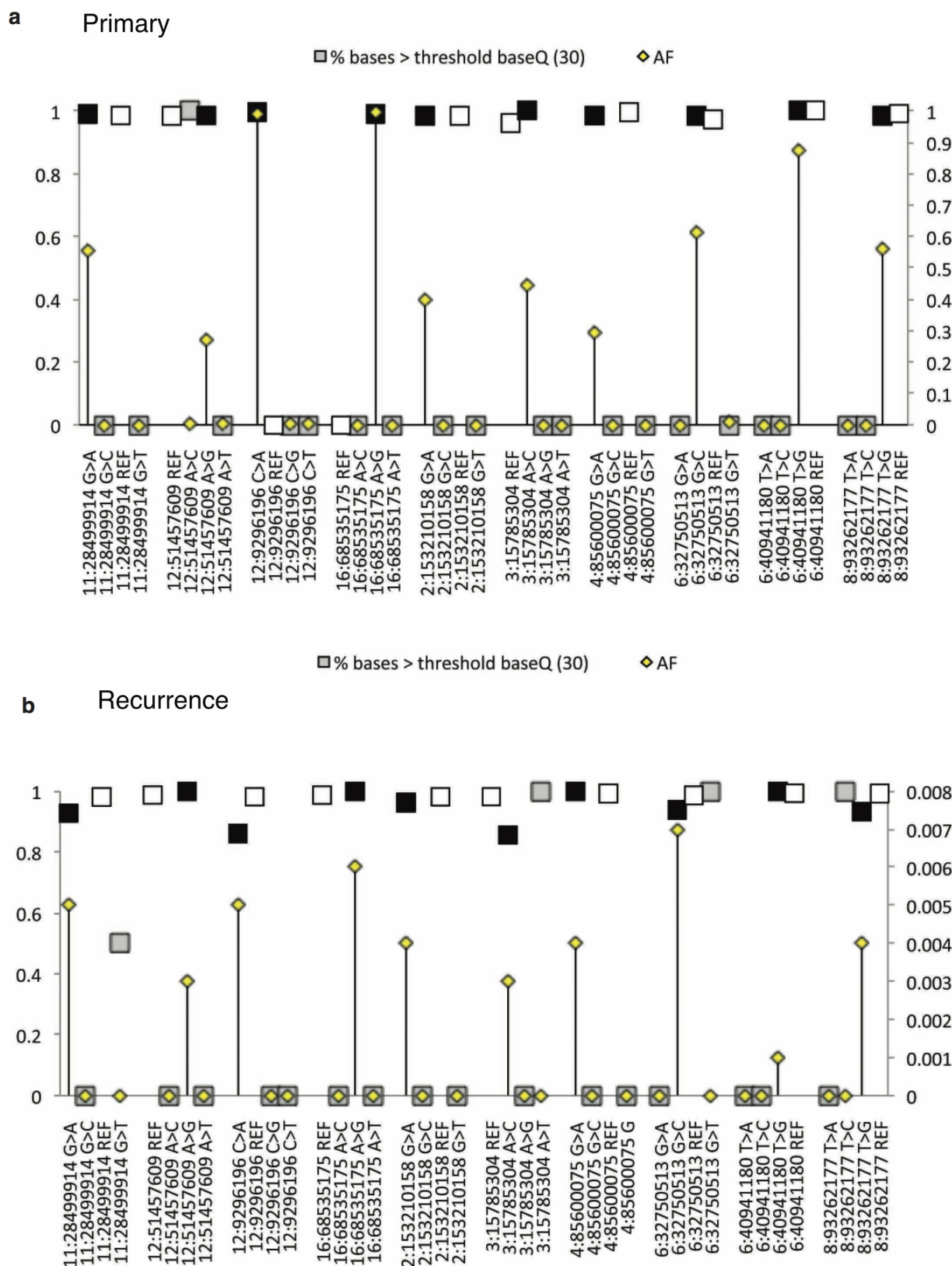
Extended Data Figure 6 | Clonal lineage evolution post-therapy in human tumours. **a**, Subpopulations of cells in each primary and recurrent tumour were identified using the EXPANDS algorithm, based on somatic SNVs and copy number gains and losses in each sample. Each subpopulation is thus distinguished by (1) a unique combination of somatic aberrations, which are (2) present in a particular subset of cells. Phylogenetic relationships between the primary (lowercase red letter labels) and recurrent (uppercase blue letter labels) tumour subpopulations indicate that in a majority of cases the recurrent tumour lineages are derived from only one lineage in the primary tumour, while only a small proportion of recurrent tumours had a more intermediate similarity to the primary tumour. **b**, The Shannon Index (SI) of each tumour is calculated using the cellular prevalence of the subpopulations defined by EXPANDS. Increasing values between the primary versus recurrent compartments indicate an increase in tumour heterogeneity (two tailed,

paired *t*-test; *P* value = 0.029, Black lines show the medians; white lines represent individual data points; polygons represent the estimated density of the data). **c**, Clonal evolution between therapy-naïve and matched recurrent tumours was assayed through ultra-deep sequencing ($>1,500\times$) of somatic mutations, and analysed using PyClone. Cellular frequencies of clones (*y* axis) are scaled by the number of mutations in each clone. Ancestral high-frequency clones present in both compartments indicate a common cell of origin in every case. Lower-frequency mutation clusters in the primary tumour indicate clones that subsequently expand to dominance in the recurrent tumour (blue lines). Higher frequency clusters in the primary tumour that are absent or extremely subclonal at the time of recurrence (red lines) indicate therapy-sensitive clones. The number of mutations studied that support each type of event are indicated in the inset box.



Extended Data Figure 7 | Subclonal expansion of rare (<5%) SNVs in the primary tumour to clonal dominance in the recurrent compartment. **a**, Deep amplicon sequencing was used to profile 20 patients with clonal SNVs restricted to their recurrent tumours as determined by 30× WGS data. Many ‘recurrence specific’ SNVs (blue) were found in a very minor subclone (<5%) of the primary tumour (red) when studied by deep amplicon sequencing. Clonal SNVs (allele frequency >15%) in recurrent tumours that had >1 read supporting an alternate base in the primary tumour are shown by patient. In each case, the number of events with evidence for expansion from a clone present at <5% is shown as a proportion (red bar) of the total number of considered events (blue bar). **b**, Evidence for clonal expansion at recurrence of clones present at <5% in the untreated tumour was observed in 16/20 patients, indicating

that clonal selection is common after therapy for medulloblastoma. The extent of clonal selection (blue > red) varies across medulloblastoma cases, with prominent clonal selection in some cases (MB-REC-30), and more extreme divergence in others (MB-REC-23). **c**, **d**, Deep amplicon sequencing of clonal SNVs from both a first recurrence (dark blue), and a subsequent second recurrence (light blue) of patient MB-REC-31 reveals that clonal SNVs present at recurrence but absent from a 30× WGS profile of the untreated tumour (red) were indeed present at very low prevalence (~1/1,000) in the primary sample, indicating striking clonal expansion after initial treatment of the untreated tumour (**c**; AF, allelic frequency; NA, not available). This is illustrated in panel **d**, which depicts the allelic frequency of a very low-prevalence *PIK3CA* mutation in the primary tumour that reaches clonal levels post-therapy.



Extended Data Figure 8 | Base quality assessment of reference and alternate alleles at SNVs with clonal or rare allelic frequencies.

a, b, To determine whether low-frequency (<5%) base calls were SNVs or sequencing errors, we analysed the distribution of base quality (baseQ) values for each alternate base called. This plot shows the allele frequencies (AF; secondary y axis) and the proportion of supporting reads with baseQ values >30 (primary y axis) for a subset of SNVs in the recurrent tumour of MB-REC-03 (x axis). At all positions, and in both the recurrent (a), and primary (b), tumours, we observe a high proportion (~100%) of reads

with baseQ >30 at both the mutant (black square) and wild-type allele (white squares). Grey squares indicate alleles categorized as sequencing errors. Errors have low allelic frequencies (in many cases are just one read) and a much smaller proportion of reads with baseQ values >30. In the primary tumour, the baseQ and AF values match the pattern observed in the recurrent tumour, indicating that these calls represent true SNVs present at very low frequencies. Sequencing errors in the primary sample have the same base distribution as sequencing errors in the recurrent tumour sample.

SHARED_G4
SHARED_G3
REFERENCE_G4
REFERENCE_G3
REFERENCE_G4
REFERENCE_G3
NAVETUM_G4
NAVETUM_G3
NAVETUM_WNT
NAVETUM_WNT

116657125123611

116657125123611

116657125123611

121171211121

121171211121

11222

121141201121

2

21920594113013516

1671849092911514

11211111

11113

11113

410441661318034104571436

410240661317534106551334

4833456131532814411235

77933541213126135371032

3371935665166320623

48333561315128115401235

4763053121332710536927

135916476105419516

25712449673116412

2

8.40e-111388

7.01e-111382

5.06e-12204

4.54e-12203

4.68e-0330

2.20e-15145

1.99e-024

1.44e-021331

2.77e-021224

3.56e-026

1.42e-023

1.42e-023

1.02e-026794

6.01e-036533

2.07e-035465

6.90e-044828

1.98e-022476

9.73e-035406

4.47e-024656

9.74e-031986

1.44e-021400

1.99e-024

GO:0022610

GO:0007155

GO:0098609

GO:0098742

GO:0016339

GO:0007156

GO:0021938

GO:0048699

GO:0030182

GO:0002309

GO:0071688

GO:0030241

GO:0032501

GO:0044707

GO:0032502

GO:0048856

GO:0009653

GO:0044767

GO:0007275

GO:0050793

GO:0045595

GO:0007418

biological adhesion (1)

cell adhesion (2)

cell-cell adhesion (3)

cell-cell adhesion via plasma-membrane adhesion molecules (4)

calcium-dependent cell-cell adhesion via plasma membrane cell adhesion mole... (5)

homophilic cell adhesion via plasma membrane adhesion molecules (5)

smoothered signaling pathway involved in regulation of cerebellar granule c... (1)

generation of neurons (1)

neuron differentiation (2)

T cell proliferation involved in immune response (1)

striated muscle myosin thick filament assembly (1)

skeletal muscle myosin thick filament assembly (1)

multicellular organismal process (1)

single-multicellular organism process (2)

developmental process (1)

anatomical structure development (2)

anatomical structure morphogenesis (2)

single-organism developmental process (1)

multicellular organismal development (2)

regulation of developmental process (1)

regulation of cell differentiation (2)

ventral midline development (1)

term name and depth in group

21413240520956

1238236410622

81322615211

81322615211

13313182866401467

16412144514835

93662338512

279122337496310

259102316468310

412549761220341184722241

1943045913322125451328

1.52e-04993

9.31e-05573

2.48e-03407

2.08e-04358

1.43e-041607

5.36e-05812

3.62e-02388

2.08e-021408

1.56e-021353

2.08e-038905

1.83e-024892

GO:0097458

GO:0036477

GO:0044297

GO:0043025

GO:0042995

GO:0043005

GO:0030425

GO:0031226

GO:0005887

GO:0016020

GO:0071944

neuron part (1)

somatodendritic compartment (2)

cell body (1)

neuronal cell body (2)

cell projection (1)

neuron projection (2)

dendrite (3)

intrinsic component of plasma membrane (1)

integral component of plasma membrane (1)

membrane (1)

cell periphery (1)

term name and depth in group

29371274341226

125

4943148168351531336

9.17e-05665

2.15e-0214

5.00e-036103

GO:0005509

GO:0052742

GO:0043167

calcium ion binding (1)

phosphatidylinositol kinase activity (1)

ion binding (1)

term name and depth in group

24

2

2

11111111

4.98e-024

4.99e-023

4.99e-023

4.97e-021

CORUM:2383

CORUM:5709

CORUM:631

CORUM:2681

ITGA5-ITGB1-FN1-TGM2 complex (1)

ArgBP2a-CBL-PTK2B complex (1)

Tacc1-chTOG-AuroraA complex (1)

p53 homotetramer complex (1)

term name and depth in group

63101361

6372474

1.19e-02214

4.87e-02141

HP:0003468

HP:0000692

Abnormality of the vertebrae (1)

Misalignment of teeth (1)

term name and depth in group

115510114410613

11214161

231615511

2115161

222615512

12116132

3.53e-02399

4.24e-0296

1.95e-0274

2.00e-0286

4.31e-0283

3.97e-0282

KEGG:05200

KEGG:04972

KEGG:05412

KEGG:04911

KEGG:05410

KEGG:04070

Pathways in cancer (1)

Pancreatic secretion (1)

Arrhythmogenic right ventricular cardiomyopathy (ARVC) (1)

Insulin secretion (1)

Hypertrophic cardiomyopathy (HCM) (1)

Phosphatidylinositol signaling system (1)

term name and depth in group

14272611354

103227114422

10641515276

12341311116

2.59e-02608

1.71e-02562

3.50e-02624

1.31e-02533

MI:hsa-miR-302d

MI:mo-miR-352

MI:hsa-miR-302d*

MI:hsa-miR-302c*

MI:hsa-miR-302d (1)

MI:mo-miR-352 (1)

MI:hsa-miR-302d* (1)

MI:hsa-miR-302c* (1)

term name and depth in group

46431211

14

14

24121

422726137

1261

51214211

5112131111

26111

3.54e-0325

4.51e-0212

1.98e-0210

2.41e-027

4.86e-02288

4.28e-0234

6.65e-03118

3.23e-03110

4.28e-0234

REAC:977068

REAC:112409

REAC:110056

REAC:5635851

REAC:112316

REAC:1660499

REAC:194138

REAC:4420097

REAC:5218921

Termination of O-glycan biosynthesis (1)

ERK activation (1)

ERK1 activation (2)

GLI proteins bind promoters of Hh responsive genes to promote transcription (1)

Neuronal System (1)

Synthesis of PIPs at the plasma membrane (1)

Signaling by VEGF (1)

VEGFA-VEGFR2 Pathway (2)

VEGFR2 mediated cell proliferation (3)

term name and depth in group

1121642

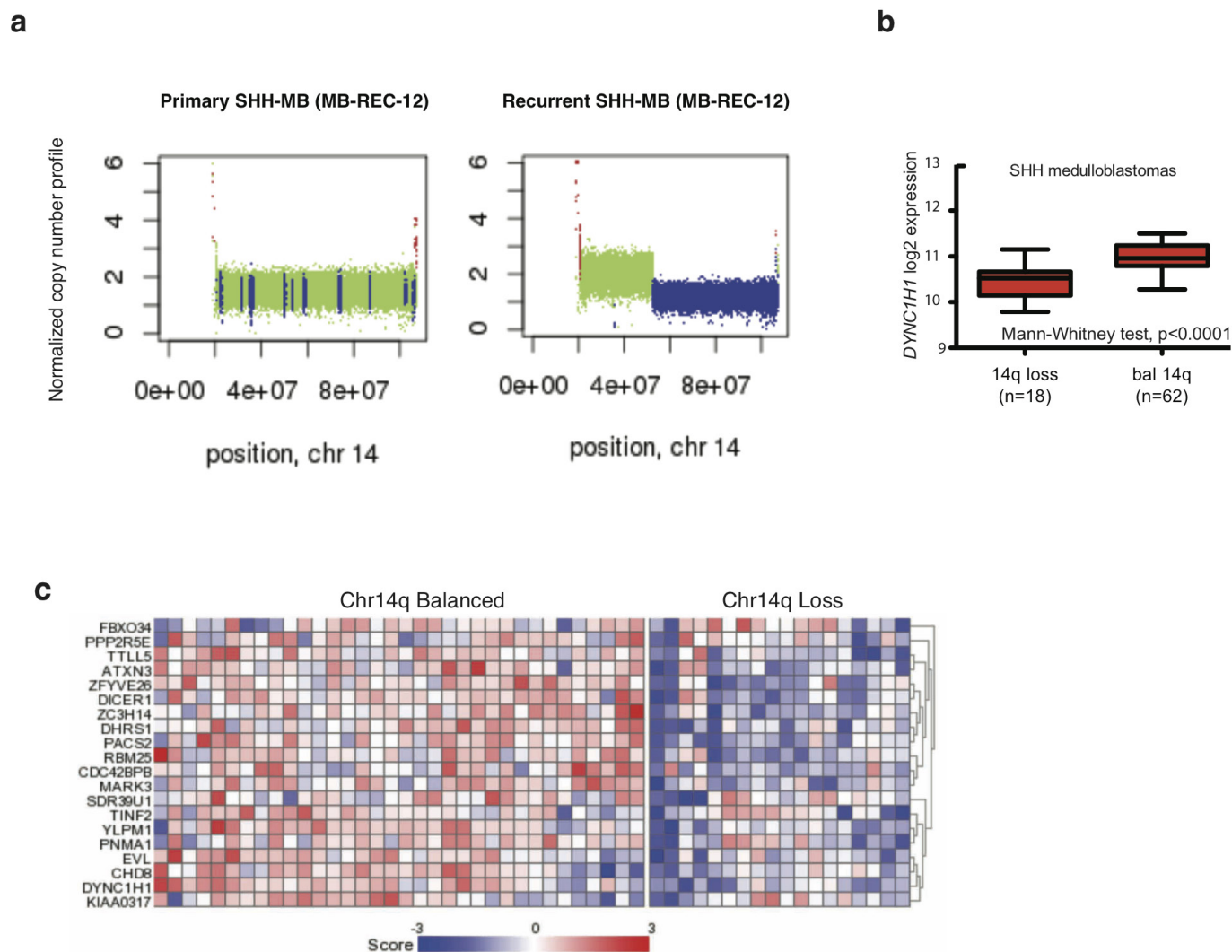
4.74e-02304

TF:M00203_3

Factor: GATA-X; motif: NGATAAGNMNN; match class: 3 (1)

Extended Data Figure 9 | Pathway enrichment results for genes recurrently aberrant in the primary tumour or recurrent tumour cohorts. Pathway enrichment analysis of gene lists derived from the

integrative analysis of CNVs (gain or loss of 2 or more copies), SNVs, indels, and structural variants specific to the primary or recurrent tumours of each patient was performed using g:Cocoa.



Extended Data Figure 10 | Genetic events in recurrent human medulloblastoma converge on specific signalling pathways. a, Copy number profile of MB-REC-12 therapy-naïve (WT, green, left panel) and recurrent tumour (loss, blue, lower panel), showing recurrence-specific loss of chr14q. **b,** *DYNC1H1* expression is reduced in Shh patients with chr14q loss ($n = 18/80$, Mann-Whitney test, $P < 0.0001$, centre lines

show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles). **c,** Expression of the chr14 signature genes discriminating between chr14q balanced ($n = 34$) and chr14q loss ($n = 18$) in the Boston cohort of Shh medulloblastoma samples.

Codon influence on protein expression in *E. coli* correlates with mRNA levels

Grégory Boël^{1,2}, Reka Letso^{1*}, Helen Neely^{1*}, W. Nicholson Price^{1†*}, Kam-Ho Wong¹, Min Su¹, Jon D. Luff¹, Mayank Valecha¹, John K. Everett³, Thomas B. Acton³, Rong Xiao³, Gaetano T. Montelione^{3,4}, Daniel P. Aalberts⁵ & John F. Hunt¹

Degeneracy in the genetic code, which enables a single protein to be encoded by a multitude of synonymous gene sequences, has an important role in regulating protein expression, but substantial uncertainty exists concerning the details of this phenomenon. Here we analyse the sequence features influencing protein expression levels in 6,348 experiments using bacteriophage T7 polymerase to synthesize messenger RNA in *Escherichia coli*. Logistic regression yields a new codon-influence metric that correlates only weakly with genomic codon-usage frequency, but strongly with global physiological protein concentrations and also mRNA concentrations and lifetimes *in vivo*. Overall, the codon content influences protein expression more strongly than mRNA-folding parameters, although the latter dominate in the initial ~16 codons. Genes redesigned based on our analyses are transcribed with unaltered efficiency but translated with higher efficiency *in vitro*. The less efficiently translated native sequences show greatly reduced mRNA levels *in vivo*. Our results suggest that codon content modulates a kinetic competition between protein elongation and mRNA degradation that is a central feature of the physiology and also possibly the regulation of translation in *E. coli*.

Degenerate encoding of 20 amino acids by 61 triplet nucleotide codons enables the same protein sequence to be synthesized by a vast number of synonymous mRNAs. While variations between synonymous sequences have an important role in regulating protein expression in organisms from *E. coli*^{1–7} to humans⁸, many details remain unclear^{1,3–5,9–13}, as summarized in the Supplementary Information. Uncertainty exists concerning the influence of synonymous codons on translation efficiency^{1,3,5,9,10,12–17}, the mechanistic basis of such effects^{4,9–11,15–31}, and their relationship to mRNA-folding effects^{3–7,16,32}. Much of the codon-usage literature focuses on inefficient translation of a set of rare codons in *E. coli*^{1,5,7,14,19,21}, especially the AUA codon for the Ile residue^{18,20} and the AGA, AGG and CGG codons for the Arg residue^{1,6,27}. On this basis, it is widely assumed that genomic codon-usage frequency, which parallels transfer RNA levels^{19,23}, tracks the translation-elongation rate, and that infrequently used codons are translated inefficiently^{9,14,16,18,19,21,24,25,27,28}. However, this assumption has been challenged by ribosome-profiling studies that concluded that the net translation-elongation rate is generally constant, irrespective of codon usage^{10,11,33}.

Investigations of the influence of mRNA sequence on protein expression are complicated by the fact that synonymous sequence changes simultaneously influence multiple parameters including codon identity, codon homogeneity and mRNA folding, as well as other local and global sequence features ranging from codon-pair effects^{34,35} to overall A/U/C/G content. Most previous studies have focused on individual parameters or pairs of parameters in a local mRNA region^{1–3,5–7,11,13}. To address this limitation, we performed statistical analyses of a large-scale protein expression data set, focusing on simultaneous evaluation of the influence of a wide variety of local and global mRNA sequence properties, and we tested the resulting mechanistic inferences using biochemical experiments. These studies provide insight into the influence of mRNA sequence features on protein expression in *E. coli*,

including a new codon-influence metric with notable differences from previously described metrics. This metric correlates strongly with physiological mRNA levels and lifetimes *in vivo*, suggesting the dynamics of the ribosomal elongation cycle exert a considerable influence on mRNA degradation that contributes to the biological effects of variations in synonymous codon usage.

Large-scale protein expression data set

We evaluated the expression of 6,348 genes from diverse phylogenetic sources (Extended Data Fig. 1), which provided broad sampling of codon space due to variations in codon-usage frequency in the source organisms. To minimize effects from coupling of translation to transcription by *E. coli* RNA polymerase^{36,37}, the genes were transcribed by bacteriophage T7 RNA polymerase from a pET plasmid³⁸. Protein expression³⁸ was induced overnight in defined medium at 17°C in *E. coli* BL21(DE3) cells containing pMGK, a compatible plasmid carrying a copy of the *argU* gene encoding the tRNA cognate to the AGA codon for Arg (refs 1, 27). The analysed proteins, selected to have less than 60% pairwise sequence identity, were expressed with a C-terminal LEHHHHHHH affinity tag that was omitted from computational analyses. On the basis of the visual inspection of whole-cell lysates in Coomassie-blue-stained SDS-PAGE gels, we scored the protein expression level on an integer scale from 0 (none) to 5 (highest) from two isolates of each plasmid, which rarely varied by more than ±1 (see supplementary fig. 1 in ref. 39). In the analysed data set, 28% and 31% of the proteins have scores of 0 and 5, respectively, while 41% have intermediate scores.

We evaluated distributions of a variety of mRNA sequence parameters, which revealed many differences between genes giving high versus low protein expression (Fig. 1 and Extended Data Fig. 2). We examined histograms of the parameter distributions for the genes giving each score (Fig. 1a–d, f, g, i and Extended Data

¹Department of Biological Sciences and Northeast Structural Genomics Consortium, 702 Fairchild Center, MC2434, Columbia University, New York, New York 10027, USA. ²CNRS UMR8261, Institut de Biologie Physico-Chimique, 13-rue Pierre et Marie Curie, 75005 Paris, France. ³Department of Molecular Biology and Biochemistry and Northeast Structural Genomics Consortium, Center for Advanced Biotechnology and Medicine, Rutgers, the State University of New Jersey, Piscataway, New Jersey 08854, USA. ⁴Department of Biochemistry, Robert Wood Johnson Medical School, Rutgers, the State University of New Jersey, Piscataway, New Jersey 08854, USA. ⁵Department of Physics, Williams College, Williamstown, Massachusetts 01267, USA. [†]Present address: WNP, University of New Hampshire School of Law, 2 White Street, Concord, New Hampshire 03301, USA.

*These authors contributed equally to this work.

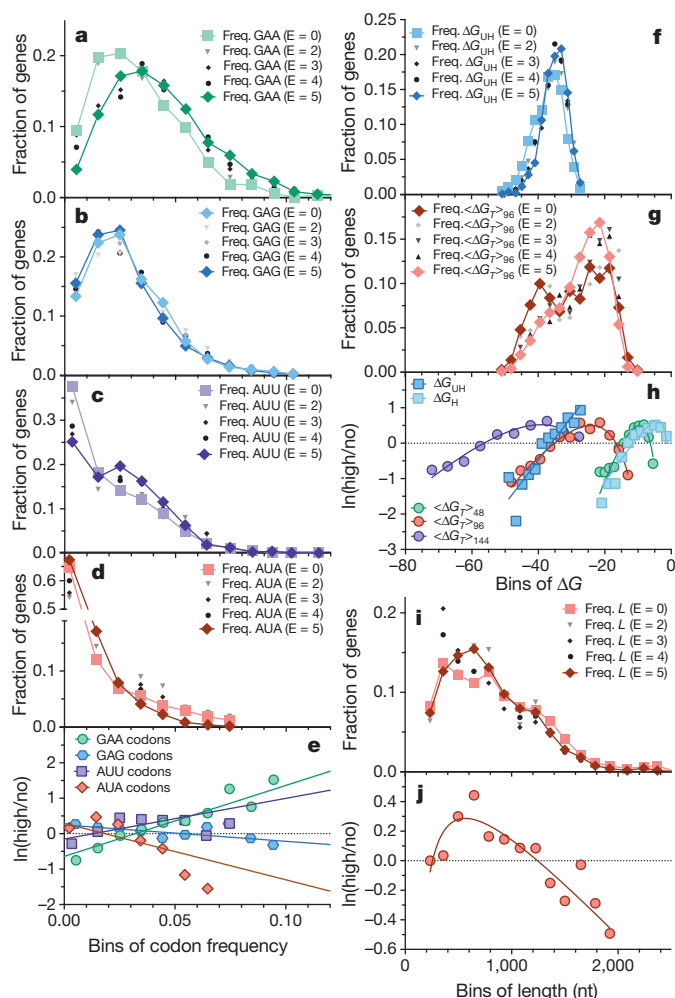


Figure 1 | Distributions of representative RNA sequence parameters in protein-expression categories in the large-scale data set. **a–d, f, g, i.** Histograms showing frequencies of GAA (**a**) and GAG (**b**) codons for Glu, AUU (**c**) and AUA (**d**) codons for Ile, partition-function free energy of folding⁴¹ of the 5'-UTR plus initial 16 codons or 'head' of each gene (ΔG_{UH} , **f**), average partition-function free energy of folding in the remainder or 'tail' of each gene in 50% overlapping windows with width w ($\langle \Delta G_T \rangle_w$, **g**), and number of nucleotides in the protein-coding sequences (**i**). Light colours, dark colours, and shades of grey show distributions in, respectively, the E = 0, E = 5, and E = 1–4 categories ($n = 6,348$ for all combined). **e, h, j.** Corresponding log-odds-ratio plots showing the logarithm of the ratio of the number of proteins in the E = 5 (E5) versus E = 0 (E0) categories ($n = 3,727$ for both combined), labeled $\ln(\text{high}/\text{no})$. Solid lines show single-variable binary generalized linear logistic regressions, that is, fitting of log-odds ratio using a combination of the first and second powers and inverse of each individual parameter. Codon regressions (solid lines in panel **e**) were performed using exclusively the first power of frequency, yielding the dark grey codon slopes in Fig. 3a. *P* values for regressions shown here other than GAA frequency are 3–45 orders of magnitude below the Bonferroni-corrected 5% confidence threshold of 2×10^{-4} (Supplementary Data File 1). nt, nucleotides.

Fig. 2a, i, k), which show roughly monotonic changes with increasing score. We also examined 'log-odds-ratio' plots of the natural logarithm of the ratio of the numbers of genes giving scores of 5 versus scores of 0 ($\ln(\text{high}/\text{no})$) as a function of each parameter value (Fig. 1e, h, j and Extended Data Fig. 2b–h, j, l–o), which provide a graphical summary of the trends, as well as guidance for mathematical modelling of the relationship between mRNA sequence parameters and protein expression. As described in the Supplementary Information, there are strong correlations between many individual mRNA sequence

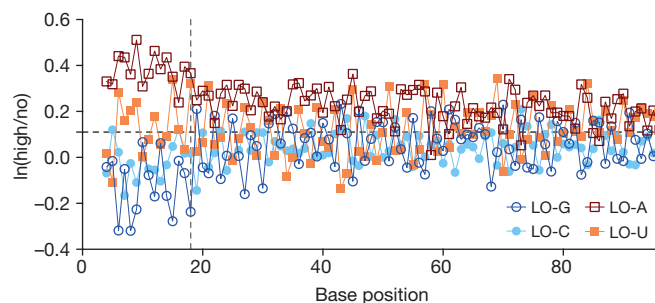


Figure 2 | Log-odds ratio for E5 versus E0 expression categories for proteins encoded by each nucleotide base at positions 4–96. The vertical grey dotted line approximates the end of the region protected by the ribosome in the 70S initiation complex. Position 1 is A in the AUG initiation codon. LO, log-odds-ratio.

parameters and the protein expression levels (Fig. 1 and Extended Data Fig. 2), but they reflect both direct effects and indirect effects caused by parameter cross-correlations (Extended Data Fig. 3). Therefore, simultaneous multi-parameter modelling is required to estimate the actual influence of individual sequence parameters.

The first 18 nucleotides in the coding sequence, which are physically protected by the ribosome in the 70S initiation complex⁴⁰, have a strong influence on expression (Fig. 2). In this region, G reduces and A increases the probability of high expression^{6,7}, whereas C and U have intermediate effects (Fig. 2). This rank-order matches the probability of base-pairing in ensembles of RNA structures (D.P.A., manuscript in preparation), suggesting that the trend reflects a requirement for bases in this region to be unpaired for efficient ribosome docking⁶. (The approximate periodicity of 3 in Fig. 2 comes from parameter cross-correlations in A/T-rich genes; see later.)

Logistic-regression analysis and modelling

We analysed the influence of mRNA sequence parameters on protein expression in our large-scale data set using logistic regression, which uses a generalized linear model to quantify the influence of continuous variables on binary or ordinal results. Binary modelling assumed that the log-odds-ratio for 5 versus 0 scores increases linearly with the value of some function of a continuous variable (for example, codon frequency), whereas ordinal modelling assumed the logs-odds-ratio between successive integer scores increases in the same manner. The solid lines in Fig. 1e show the most probable slopes for a linear relationship between codon frequencies and the log-odds-ratio of 5 versus 0 expression scores, which is the simplest form of binary logistic regression modelling. This simple linear model accurately describes the positive correlation of GAA (green in Fig. 1e). It is less accurate describing the negative correlation of AUA. Logistic regression can be performed using more complex mathematical functions of the continuous variable. Nonetheless, 'codon slopes' from linear logistic-regressions provide a useful metric to quantify the influence of individual codons on protein expression.

We conducted such single-variable analyses on all 61 non-stop codons using binary or ordinal linear logistic regression (dark and light grey, respectively, in Fig. 3a). Relatively uniform variance in codon frequencies (Extended Data Fig. 4a) enables parameters for all codons to be determined with similar precision. Binary and ordinal regressions yield equivalent codon slopes, reinforcing our inference from parameter histograms (Fig. 1a–d) that codon content has a roughly monotonic influence on expression. The equivalence of results comparing proteins with 0 versus 5 scores to results including proteins with intermediate scores suggests features that partially attenuate expression also sometimes completely stop it, consistent with the data presented below linking features that attenuate translation to mRNA degradation.

Single-parameter logistic regressions (Fig. 3a grey symbols and Extended Data Fig. 4b) show codons ending in A/U are enriched in genes

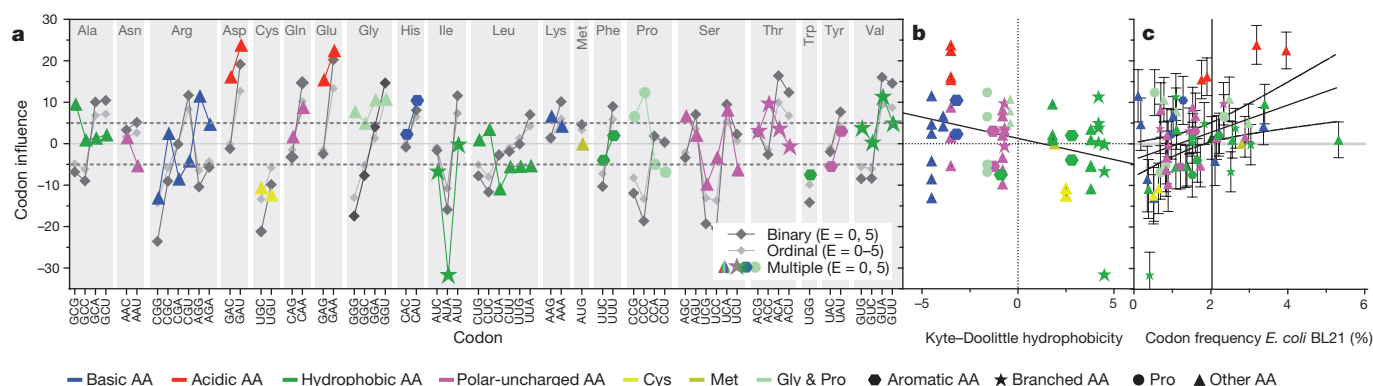


Figure 3 | Codon influence on protein expression in the large-scale data set. **a**, Slopes for every non-stop codon from single parameter binary logistic regression analyses of the E = 0 versus E = 5 categories (dark grey, $n = 3,727$ combined), single parameter ordinal logistic regression analyses of the E = 0–5 categories (light grey, $n = 6,348$ combined), and simultaneous multi-parameter binary logistic regression analysis of the

giving the highest expression, whereas synonymous codons ending in G/C are depleted. These results provide guidance for engineering genes that enhance protein expression by emulating the properties of our best-expressed genes, a strategy we show later to be successful. However, these analyses do not provide reliable information on the influence of individual codons because the frequencies of codons ending in A and U are mutually correlated in genes in our data set (Extended Data Fig. 3a–c), owing at least in part to variations in the A/T frequency in the source organisms. Many parameters that vary systematically with expression level are mutually correlated (Extended Data Fig. 3). Parameters that do not directly influence outcome can appear influential in single-parameter regressions when their values are correlated with directly influential parameters.

Therefore, to dissect the mechanistic contributions of the parameters, we performed multi-parameter binary logistic-regression modelling, which simultaneously analyses the influence of all parameters, although reliability in quantifying the influence of correlated parameters depends on the extent to which they vary independently in the data set. Our final model (model M in Extended Data Table 1) combines explanatory variables (Figs 1, 2 and Extended Data Fig. 2) after eliminating those redundant with correlated variables. The logarithm of the odds of observing a 5 versus 0 expression score is given by:

$$\theta = 2.0 + 0.054\Delta G_{UH} - 1.5I + 6.6a_H - 6.3a_H^2 - 1.8g_H^2 + 0.80u_{3H} + 0.86 \sum_c \beta_c f_c + 0.078s_{7-16} + 0.063s_{17-32} - 1.8d_{AUA} - 16r - 0.0012L - 520/L$$

in which ΔG_{UH} is the predicted free energy of folding⁴¹ of the head plus 5'-untranslated region (UTR), I is a binary indicator that is 1 only if $\Delta G_{UH} < -39 \text{ kcal mol}^{-1}$ and the GC content of codons 2–6 is greater than 62%, a_H and g_H are A and G frequencies in codons 2–6, u_{3H} is the U frequency at the third position in codons 2–6, β_c and f_c are the slopes (coloured symbols in Fig. 3a) and frequencies of each non-termination codon, s_{7-16} and s_{17-32} are the mean slopes for codons 7–16 and 17–32, d_{AUA} is a binary variable that is 1 only if there is at least one AUA–AUA di-codon, r is the amino acid repetition rate, and L is the sequence length.

Calculating loss in predictive power when model terms are omitted gives the best estimate of the influence of sequence parameters (Fig. 4). The influence of the head is captured by the folding energy and base-composition terms, which probably reflect accessibility of the translation initiation site for ribosome docking^{6,7,40}, together with s_{7-16} . The influence of the tail is captured by s_{17-32} together with the global terms (overall codon content, d_{AUA} , r and L). Our model indicates the influential mRNA-folding effects are restricted to the head

E = 0 versus E = 5 categories (model M in Extended Data Table 1, coloured symbols). Shapes and colours encode the structures and qualitative chemical characteristics of the side chains, respectively. AA, amino acids. **b**, **c**, Codon slopes from the multi-parameter binary logistic regression plotted against Kyte–Doolittle amino-acid hydrophobicity (**b**) or *E. coli* BL21 codon-usage frequency (**c**).

and are somewhat weaker than codon effects (Fig. 4b). The influence of individual codons is ~ 3 times stronger near the start of the gene and declines to a uniform level after codon ~ 32 (Extended Data Fig. 5), roughly matching the number of residues filling the ribosomal exit channel⁴². However, total codon content in the tail is on average ~ 5 times more influential than in the head because the tail is substantially longer. In-frame codon models are superior to out-of-frame and expanded models including adjacent 3' or 5' bases (Extended Data Table 1a and Methods.) Calculations also show that the correlation between expression and the average predicted mRNA folding energy in the tail ($< G_T >_{96}$) is attributable to its correlation with codon slopes and the amino acid repetition rate (Extended Data Fig. 3d, e).

The codon slopes from our multi-parameter logistic-regression model (coloured symbols in Fig. 3a) provide a new metric quantifying the influence of each codon on translation efficiency in *E. coli*. While some features of this codon-influence metric match previous conclusions, broad trends do not. The AUA codon for Ile, which is decoded by a non-cognate tRNA^{18,20}, has the strongest expression-attenuating effect, and this effect is enhanced for adjacent pairs of AUAs (Extended Data Fig. 2f). The approximately neutral influence of the other codons for Ile indicates the expression-attenuating effect of AUA

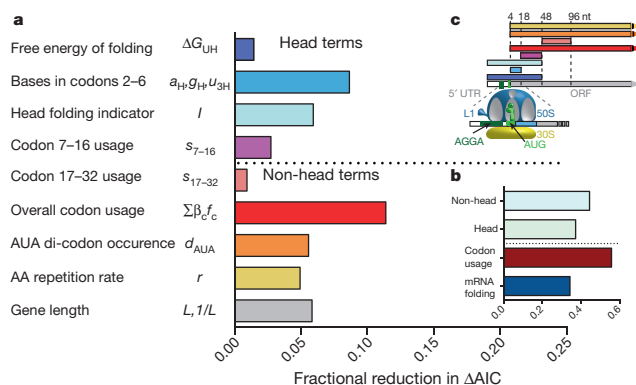


Figure 4 | Contributions of physicochemical factors and regions of the coding sequence to protein expression level. **a**, **b**, Bar graphs showing the fractional reduction in the magnitude of ΔAIC when omitting individual (**a**) or combinations (**b**) of terms before re-optimizing the remaining terms in model M (Extended Data Table 1). ΔAIC , the change in the Akaike information criterion, quantifies the predictive power of the model compared to random expectation. **c**, Schematic illustrating the region of the protein-coding sequence included when calculating the term represented by the same colour. Numbering starts at the first nucleotide in the start codon. Terms related to mRNA folding are shown in blue and cyan. Those related to codon usage are shown in red, orange, yellow and magenta.

is not attributable to amino acid structure. Similarly, the CGG and CGA codons for Arg have strong expression-attenuating effects, while the four synonymous codons have weaker effects varying in direction. Among rare codons emphasized in the past to be deleterious^{1,18,19,24}, four attenuate expression in our data set (those cited above and CUA for Leu), whereas the other four do not (AGA and AGG for Arg, GGA for Gly, and CCC for Pro). The apparent influence of AGA and AGG, which have the lowest frequencies in *E. coli*, may be biased by overexpression in our experiments of the ArgU tRNA cognate to AGA^{1,27}. The next three least frequent codons attenuate expression to widely varying extents, and codons with slightly higher frequencies do not (Fig. 3c and Extended Data Fig. 4d). There is no significant correlation between the frequencies of the remaining 56 non-stop codons and their influence on expression (Fig. 3c and Extended Data Fig. 4d). Similarly, there is at most a weak correlation between our codon slopes and the codon adaptation index¹⁴, the codon sensitivity²⁴, the tRNA adaptation index¹⁶, or the estimated cognate tRNA concentrations²³ (Extended Data Fig. 4e–h, respectively).

The most strongly expression-enhancing codons in Fig. 3a encode amino acids with side chains that can act as general base catalysts (Glu, Asp and His). Their codons ending A/U have stronger enhancing effects than synonymous codons ending G/C, indicating that codon structure may modulate their translation efficiency. However, plotting codon slopes against amino acid hydrophobicity reveals a strong correlation (Fig. 3b), with charged residues having higher slopes than polar or hydrophobic residues, suggesting that translation efficiency varies systematically with amino acid structure. Analysing the slopes according to the nucleotide at each codon position reveals trends (Extended Data Fig. 4c) that are likely to reflect conservation of physicochemical properties of amino acids among codons with the same bases at positions 1–2 rather than differences in translation efficiency attributable to base content.

Design and testing of efficiently translated genes

To test the predictive value of these analyses, we evaluated expression of synthetic genes designed using two methods emulating the codon-usage and mRNA-folding properties of genes giving the highest protein levels

in our data set (Fig. 5 and Extended Data Figs 6 and 7 and Extended Data Table 2). The ‘six amino acid’ (6AA) method substitutes all Arg, Asp, Glu, Gln, His and Ile codons with the synonymous codon having the highest single-variable logistic regression slope (dark grey in Fig. 3a). The resulting mRNAs are enriched in codons ending A/U, which give weaker folding energies than codons ending G/C. These mRNAs tend to have folding and other properties matching the genes giving the highest expression in our data set, providing a concrete example of the parameter cross-correlations shown in Extended Data Fig. 3a–c. The ‘31 codon folding optimization’ (31C-FO) method explicitly optimizes mRNA folding using just 31 codons with the highest single-variable logistic regression slopes for each amino acid; the folding energy⁴¹ in the head (ΔG_{UH}) was maximized (that is, minimizing folding stability), whereas that in the tail ($<\Delta G_T>_{48}$) was adjusted to be near -10 kcal mol⁻¹. Some genes were engineered in the head but not the tail, or vice versa, to evaluate our inferences concerning their relative contributions.

We synthesized 31C-FO genes for five proteins (Fig. 5 and Extended Data Fig. 6), including one from *E. coli*, that were poorly expressed using their native genes, and for 17 previously uncharacterized proteins (Extended Data Fig. 7a). These codon-optimized genes give uniformly high expression (scores of 4–5 for 18 out of 19 proteins <450 amino acids in length). Equivalent improvements were observed for three pairs of native and 31C-FO genes transcribed at physiological levels from pBAD vectors by *E. coli* RNA polymerase (Extended Data Fig. 6e, f). Although some 31C-FO genes yield insoluble proteins using our standard expression protocol, they consistently yield high levels of soluble protein (Extended Data Fig. 7b) when fused to the carboxy terminus of *E. coli* maltose-binding protein, a solubility-enhancing tag.

We retained native head sequences and optimized exclusively the tails of four genes using the 6AA method (WT_H/6AA_T in Fig. 5b and Extended Data Fig. 6), which consistently increased expression, albeit to a varying extent. We also tested the relative influence of codon usage versus mRNA folding in the head by constructing genes with identical tails but different codon-optimized 31C heads with folding energies either optimized (31C-FO_H with ΔG_{UH} maximized) or

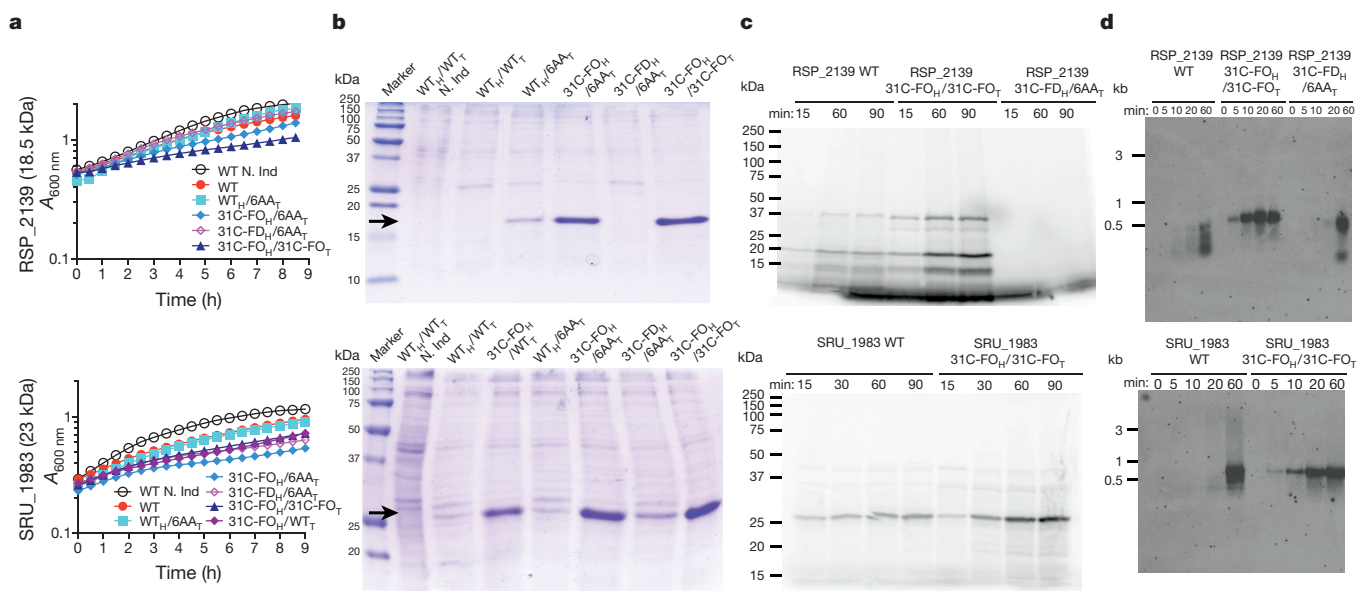


Figure 5 | Analyses of synthetic genes designed to enhance protein expression. Synonymous variants of inefficiently translated native (wild type, WT) genes were redesigned in the head or tail or both using the 6AA, 31-codon folding optimized (31C-FO) or 31-codon folding deoptimized (31C-FD) methods. The type of sequence in the head (H) and tail (T) is indicated separately. N. Ind. denotes non-induced control. **a**, Growth curves at room temperature after induction at time zero in *E. coli* BL21(DE3). **b**, Coomassie-blue-stained SDS-PAGE gels of *in vivo* expression in cells induced overnight at 17°C, with loads normalized

to final $A_{600\text{ nm}}$. Black arrows indicate target proteins. kDa, kilodaltons. **c**, Autoradiographs of SDS-PAGE gels of *in vitro* translation reactions in the presence of [³⁵S]methionine using fully purified components to translate an equal amount of purified mRNA transcribed *in vitro* by T7 RNA polymerase. Higher molecular mass bands represent SDS-resistant oligomers. **d**, Northern blots of equal amounts of total RNA isolated at the indicated times after induction, hybridized with a probe matching the 5'-UTR. kb, kilobases. All experiments yielded equivalent results when performed at least twice.

deoptimized (31C-FD_H with ΔG_{UH} minimized). These results (Fig. 5b and Supplementary Information) demonstrate that folding and codon usage in the head and codon usage in the tail all influence expression, consistent with our computational inferences (Fig. 4).

Biochemical analyses of optimized synthetic genes

We compared the cellular growth rates (Fig. 5a and Extended Data Fig. 6a), protein expression levels (Fig. 5b and Extended Data Fig. 6b) and mRNA levels (Fig. 5d and Extended Data Fig. 6d) after induction of native versus optimized genes *in vivo* in *E. coli*. We also compared *in vitro* transcription (Extended Data Fig. 8) and translation (Fig. 5c and Extended Data Fig. 6c) reactions conducted using these genes. Inhibition of growth after induction of the APE_0230.1 gene is eliminated by codon optimization even though protein expression increases greatly (Extended Data Fig. 6a, b), suggesting that inefficient translation can cause toxicity. Native versus optimized genes are transcribed *in vitro* by T7 RNA polymerase with equivalent rates and yields (Extended Data Fig. 8). However, *in vitro* translation of the resulting mRNAs yields more protein from the optimized sequences (Fig. 5c), and translational pausing sites are different for some synonymous pairs (for example, APE_0230.1). Therefore, translation efficiency is improved by the codon-optimization methods derived from our computational analyses.

We observe lower mRNA levels *in vivo* after induction of the inefficiently translated native versus optimized genes (Fig. 5d and Extended Data Fig. 6d), suggesting that translational obstacles reduce the steady-state mRNA level. Notably, 5 min after induction, full-length mRNA is detected for all of the optimized genes, but none of the native genes. Because T7 polymerase transcribes them equivalently *in vitro* (Extended Data Fig. 8), these results suggest the inefficiently translated native mRNAs are rapidly degraded. We evaluated the physiological relevance of this inference by calculating s_{all} , the average codon slope from our multi-parameter logistic-regression modelling (coloured symbols in Fig. 3a), for all *E. coli* genes. Supporting the validity of our new codon-influence metric, this parameter derived from our expression data set correlates strongly with *in vivo* protein concentrations measured using mass spectrometry⁴³ (Fig. 6b) and with the probability of detecting a protein using this technique (Fig. 6c), which increases with concentration. Notably, s_{all} correlates as strongly with *in vivo* mRNA levels for all predicted cytoplasmic proteins (Fig. 6a, b), suggesting that codon content influences the steady-state mRNA concentration. Finally, s_{all} calculated in the proper reading frame is strongly positively correlated with the lifetimes of monocistronic mRNAs in *E. coli*⁴⁴ (Fig. 6d), consistent with our inference that mRNAs encoded by better codons tend to be degraded more slowly. Acquisition of an mRNA lifetime measurement, which is biased towards higher steady-state concentration and longer lifetime, also depends strongly on s_{all} (Fig. 6c). These global correlations suggest codon content exerts an important influence on not only translation efficiency but also on mRNA stability.

Discussion

We used simultaneous multi-parameter computational modelling of results from 6,348 independent protein expression experiments to dissect the mRNA sequence features controlling protein expression in *E. coli* (Figs 1–4), and we verified these computational inferences using biochemical methods (Fig. 5). The average value (s_{all}) of the resulting codon-influence metric correlates strongly with endogenous *in vivo* protein concentrations (Fig. 6b, c), as well as mRNA concentrations (Fig. 6a–c) and lifetimes (Fig. 6d) in *E. coli*. These global correlations could derive in part from parallel evolutionary selection for efficient transcription and translation and high mRNA stability. However, *in vivo* and *in vitro* biochemical studies we conducted on synthetic codon-optimized genes support the hypothesis that codon content directly modulates both translation efficiency (Fig. 5c and Extended Data Fig. 6c) and mRNA stability (Fig. 5d and Extended Data Figs 6d and 8) in *E. coli* and that these parameters are tightly coupled⁴⁵,

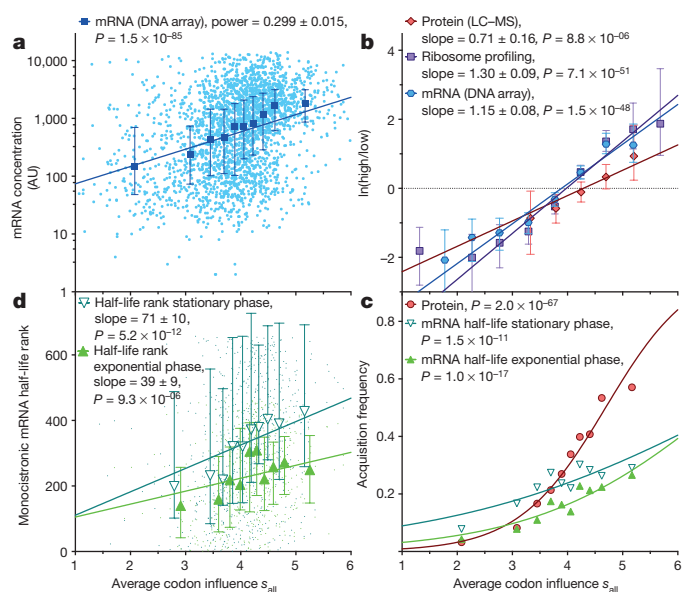


Figure 6 | Codon influence on protein expression correlates with endogenous *E. coli* protein levels and mRNA levels and lifetimes.

a, Logarithm of mRNA level of every predicted cytoplasmic protein in *E. coli* plotted versus s_{all} , the average of our codon-influence metric (coloured symbols in Fig. 3a). Cyan dots show individual genes in a microarray analysis. Blue symbols/bars show the corresponding decile plot (that is, the median, twenty-fifth and seventy-fifth percentiles in bins of s_{all} with equal populations). P value is from linear regression. **b**, Log-odds-ratio plot of predicted cytoplasmic genes/proteins in the top versus bottom 30% of the population in genome-scale *in vivo* profiles of exponentially growing cells in defined medium. Cyan, red and magenta represent data from, respectively, the microarray analysis (**a**, $n = 2,817$), a mass spectrometric analysis of protein concentration⁴³ ($n = 825$), and a deep-sequencing analysis of ribosome distribution on mRNAs¹⁰ ($n = 2,597$). Bins contain equal numbers of genes/proteins in each data set. Error bars show 95% bootstrapping confidence limits. P values are from binary logistic regression. **c**, Decile plot showing fraction of all predicted cytosolic genes/proteins acquired in the mass spectrometric analysis of protein concentration⁴³ (red) and in deep-sequencing analyses of mRNA lifetimes⁴⁴ in exponential (green) or early stationary (teal) phase in Luria Broth (LB). P values are from binary linear logistic regressions. **d**, Decile plot showing rank-order of mRNA lifetimes⁴⁴ as a function of s_{all} in the proper reading frame. P values are from Spearman correlation analysis; equivalent analyses in frames +1/+2 give P values of 0.9/0.007 and 0.6/0.4 for exponential and stationary phase data sets, respectively.

as previously shown for some individual genes^{2,36,46,47}. Recent reports suggest similar coupling in yeast^{29,30}. Several molecular mechanisms could explain this coupling^{27,36,37,47–50}, as outlined in the Supplementary Information. Modulation of mRNA stability by codon usage would enable it to influence protein expression without altering the net protein-elongation rate^{10,11}. However, our *in vitro* translation assays (Fig. 5c and Extended Data Fig. 6c) suggest that codon content also directly affects translation efficiency.

Our new codon-influence metric (Fig. 3a) has notable differences compared to previous inferences. Amino-acid identity influences protein expression efficiency, but genomic codon-usage frequency^{14,19,21} is not broadly correlated with it (Fig. 3c and Extended Data Fig. 4d). Although the third, fourth and fifth least frequent *E. coli* codons have the most deleterious influence on expression in our data set, they attenuate it to widely varying extents, and slightly more frequent codons have a neutral or expression-enhancing influence (Fig. 3c and Extended Data Fig. 4d). Codon-usage frequency correlates with the concentration of the cognate tRNA^{18–20,23}, which can influence the protein-elongation rate *in vitro*^{9,18,21,26} and protein yield *in vivo*^{1,2,27}. Indeed, ArgU tRNA was overexpressed in our experiments to promote higher expression of proteins enriched in AGA/AGG codons^{1,27}, which

may bias the influence of these codons in our data set (Fig. 3a). Further research will be required to understand the factors determining when tRNA concentration influences translation efficiency. Nonetheless, our analyses suggest ribosomal elongation dynamics^{29,35} generally exert a stronger influence on protein expression than tRNA concentration. Therefore, translational regulatory effects could operate via modification of ribosomal elongation dynamics, mediated, for example, by covalent modification of tRNAs or the ribosome²⁰. Complicating analyses of such regulatory effects^{2,36,45,46}, our results suggest they could be manifested via alterations in mRNA levels due to intimate coupling between mRNA stability and translation efficiency.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 November 2014; accepted 1 December 2015.

Published online 13 January 2016.

- Chen, G. T. & Inouye, M. Role of the AGA/AGG codons, the rarest codons in global gene expression in *Escherichia coli*. *Genes Dev.* **8**, 2641–2652 (1994).
- Deana, A., Ehrlich, R. & Reiss, C. Synonymous codon selection controls *in vivo* turnover and amount of mRNA in *Escherichia coli* *bla* and *ompA* genes. *J. Bacteriol.* **178**, 2718–2720 (1996).
- Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324**, 255–258 (2009).
- Tuller, T., Waldman, Y. Y., Kupiec, M. & Ruppin, E. Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl Acad. Sci. USA* **107**, 3645–3650 (2010).
- Goodman, D. B., Church, G. M. & Kosuri, S. Causes and effects of N-terminal codon bias in bacterial genes. *Science* **342**, 475–479 (2013).
- Castillo-Méndez, M. A., Jacinto-Loeza, E., Olivares-Trejo, J. J., Guarneros-Pena, G. & Hernandez-Sanchez, J. Adenine-containing codons enhance protein synthesis by promoting mRNA binding to ribosomal 30S subunits provided that specific tRNAs are not exhausted. *Biochimie* **94**, 662–672 (2012).
- Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z. & Bluthgen, N. Efficient translation initiation dictates codon usage at gene start. *Mol. Syst. Biol.* **9**, 675 (2013).
- Hunt, R. C., Simhadri, V. L., landoli, M., Sauna, Z. E. & Kimchi-Sarfaty, C. Exposing synonymous mutations. *Trends Genet.* **30**, 308–321 (2014).
- Spencer, P. S., Siller, E., Anderson, J. F. & Barral, J. M. Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *J. Mol. Biol.* **422**, 328–335 (2012).
- Li, G. W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157**, 624–635 (2014).
- Li, G.-W., Oh, E. & Weissman, J. S. The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484**, 538–541 (2012).
- Gingold, H. & Pilpel, Y. Determinants of translation efficiency and accuracy. *Mol. Syst. Biol.* **7**, 481 (2011).
- Cannarozzi, G. et al. A role for codon order in translation dynamics. *Cell* **141**, 355–367 (2010).
- Sharp, P. M. & Li, W. H. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**, 1281–1295 (1987).
- Ninio, J. Fine tuning of ribosomal accuracy. *FEBS Lett.* **196**, 1–4 (1986).
- Tuller, T. et al. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344–354 (2010).
- Wallace, E. W., Airolidi, E. M. & Drummond, D. A. Estimating selection on synonymous codon usage from noisy experimental data. *Mol. Biol. Evol.* **30**, 1438–1453 (2013).
- Caskey, C. T., Beaudet, A. & Nirenberg, M. RNA codons and protein synthesis. 15. Dissimilar responses of mammalian and bacterial transfer RNA fractions to messenger RNA codons. *J. Mol. Biol.* **37**, 99–118 (1968).
- Ikemura, T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.* **151**, 389–409 (1981).
- Muramatsu, T. et al. Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. *Nature* **336**, 179–181 (1988).
- Zhang, S. P., Zubay, G. & Goldman, E. Low-usage codons in *Escherichia coli*, yeast, fruit fly and primates. *Gene* **105**, 61–72 (1991).
- Bulmer, M. The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129**, 897–907 (1991).
- Dong, H., Nilsson, L. & Kurland, C. G. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J. Mol. Biol.* **260**, 649–663 (1996).
- Elf, J., Nilsson, D., Tenson, T. & Ehrenberg, M. Selective charging of tRNA isoacceptors explains patterns of codon usage. *Science* **300**, 1718–1722 (2003).
- Dittmar, K. A., Sorensen, M. A., Elf, J., Ehrenberg, M. & Pan, T. Selective charging of tRNA isoacceptors induced by amino-acid starvation. *EMBO Rep.* **6**, 151–157 (2005).
- Zhang, F., Saha, S., Shabalina, S. A. & Kashina, A. Differential arginylation of actin isoforms is regulated by coding sequence-dependent degradation. *Science* **329**, 1534–1537 (2010).
- Vivanco-Domínguez, S. et al. Protein synthesis factors (RF1, RF2, RF3, RRF, and tmRNA) and peptidyl-tRNA hydrolase rescue stalled ribosomes at sense codons. *J. Mol. Biol.* **417**, 425–439 (2012).
- Dana, A. & Tuller, T. The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res.* **42**, 9171–9181 (2014).
- Pelechano, V. & Wei, W. & Steinmetz, Lars M. Widespread co-translational RNA decay reveals ribosome dynamics. *Cell* **161**, 1400–1412 (2015).
- Presnyak, V. et al. Codon optimality is a major determinant of mRNA stability. *Cell* **160**, 1111–1124 (2015).
- Drummond, D. A. & Wilke, C. O. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134**, 341–352 (2008).
- Shakin-Eshleman, S. H. & Liebhafner, S. A. Influence of duplexes 3' to the mRNA initiation codon on the efficiency of monosome formation. *Biochemistry* **27**, 3975–3982 (1988).
- Quax, T. E. et al. Differential translation tunes uneven production of operon-encoded proteins. *Cell Rep.* **4**, 938–944 (2013).
- Letzring, D. P., Wolf, A. S., Brule, C. E. & Grayhack, E. J. Translation of CGA codon repeats in yeast involves quality control components and ribosomal protein L1. *RNA* **19**, 1208–1217 (2013).
- Ude, S. et al. Translation elongation factor EF-P alleviates ribosome stalling at polypurine stretches. *Science* **339**, 82–85 (2013).
- lost, I. & Dreyfus, M. The stability of *Escherichia coli* lacZ mRNA depends upon the simultaneity of its synthesis and translation. *EMBO J.* **14**, 3252–3261 (1995).
- lost, I., Guillerez, J. & Dreyfus, M. Bacteriophage T7 RNA polymerase travels far ahead of ribosomes *in vivo*. *J. Bacteriol.* **174**, 619–622 (1992).
- Acton, T. B. et al. Robotic cloning and protein production platform of the Northeast Structural Genomics Consortium. *Methods Enzymol.* **394**, 210–243 (2005).
- Price, W. N. et al. Large-scale experimental studies show unexpected amino acid effects on protein expression and solubility *in vivo* in *E. coli*. *Microb. Inform. Exp.* **1**, 6 (2011).
- Duval, M. et al. *Escherichia coli* ribosomal protein S1 unfolds structured mRNAs onto the ribosome for active translation initiation. *PLoS Biol.* **11**, e1001731 (2013).
- Reuter, J. S. & Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 129 (2010).
- Lu, J. & Deutsch, C. Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J. Mol. Biol.* **384**, 73–86 (2008).
- Ishihama, Y. et al. Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genomics* **9**, 102 (2008).
- Chen, H., Shiroguchi, K., Ge, H. & Xie, X. S. Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Mol. Syst. Biol.* **11**, 781 (2015).
- dos Reis, M. Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. *Nucleic Acids Res.* **31**, 6976–6985 (2003).
- Nogueira, T., de Smit, M., Graffe, M. & Springer, M. The relationship between translational control and mRNA degradation for the *Escherichia coli* threonyl-tRNA synthetase gene. *J. Mol. Biol.* **310**, 709–722 (2001).
- Richards, J., Sundermeier, T., Svetlanov, A. & Karzai, A. W. Quality control of bacterial mRNA decoding and decay. *Biochim. Biophys. Acta* **1779**, 574–582 (2008).
- Ivanova, N., Pavlov, M. Y. & Ehrenberg, M. tmRNA-induced release of messenger RNA from stalled ribosomes. *J. Mol. Biol.* **350**, 897–905 (2005).
- Shoemaker, C. J., Eyler, D. E. & Green, R. Dom34/Hbs1 promotes subunit dissociation and peptidyl-tRNA drop-off to initiate no-go decay. *Science* **330**, 369–372 (2010).
- Chadani, Y., Ono, K., Kutsukake, K. & Abo, T. *Escherichia coli* YaeJ protein mediates a novel ribosome-rescue pathway distinct from SsrA- and ArfA-mediated pathways. *Mol. Microbiol.* **80**, 772–785 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by NIGMS Protein Structure Initiative grant U54-GM094597 to the Northeast Structural Genomics Consortium to J.F.H. and G.T.M., and NIH grant GM106372 to D.P.A. We thank B. Klingenberg, R. Gonzalez, M. Gottesman and V. de Crécy-Lagard for advice.

Author Contributions T.B.A., R.X., J.K.E., J.F.H. and G.T.M. developed the protein expression platform. T.B.A., R.X. and G.T.M. generated the expression data set. W.N.P. initiated and M.S., M.V., J.D.L., G.B., J.F.H. and D.P.A. completed the computational analyses. H.N., K.-H.W. and R.X. constructed genes that R.L., K.-H.W. and G.B. used for biochemical studies. G.B., G.T.M., D.P.A. and J.F.H. designed the research and wrote the paper.

Author Information Microarray data and the pMGK sequence were deposited in the Gene Expression Omnibus and GenBank under accessions GSE73416 and KT203761, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.F.H. (jfh21@columbia.edu) or D.P.A. (aalberts@williams.edu).

METHODS

Proteins in the large-scale expression data set. The data set analysed in this paper was culled from that described in our previous paper analysing correlations between amino acid sequence and protein expression/solubility levels³⁹. In brief, proteins were selected from a wide variety of source organisms based on structural uniqueness, meaning that no sequence with greater than 30% amino acid identity had an experimentally determined structure deposited into the Protein Data Bank at the time of selection. We restricted the data set compared to that used in our earlier paper to contain only non-redundant proteins encoded by genes that do not contain any codons affected by an alternative translation table in the source organism and that were expressed with a C-terminal LEHHHHHH tag. Homologous sequences were eliminated using an iterative procedure that reduced the level of amino acid sequence identity between any pair to less than 60%, which results in a lower level of nucleic acid sequence identity. At each step, all pairs of proteins sharing at least 60% identical amino acid sequence identity were transitively grouped together into a set, and the shortest sequence was eliminated from each set before reinitiating the same set-assignment procedure on all remaining proteins. The resulting data set included 6,348 genes from 171 organisms, as detailed in the cladogram in Extended Data Fig. 1 and Supplementary Data File 2. It contained 95 endogenous *E. coli* genes, including *ycaQ* that was examined in our follow-up biochemical experiments (Extended Data Fig. 6), and 6,253 genes from heterologous sources, including 47 from mammals, 809 from archaeobacteria, and the remainder from 151 different eubacterial organisms.

Scoring of protein expression in the large-scale data set. The methods used in our large-scale protein expression experiments were described in detail previously^{38,51,52}, and they are similar to those described below for evaluation of protein expression *in vivo* except that induction was performed in 0.5-ml cultures in 96-well plates. In brief, native genes for the 6,348 proteins were cloned with a C-terminal LEHHHHHH affinity tag under the control of the bacteriophage T7 promoter in pET21, a 5.4-kb pBR322-derived plasmid harbouring an ampicillin resistance marker³⁸. Protein expression³⁸ was induced overnight at 17 °C in *E. coli* strain BL21(DE3) growing in chemically defined medium containing glucose as a carbon source. The expression strain also contained pMGK (GenBank accession number KT203761), a 5.4-kb pACYC177-derived plasmid that harbours a kanamycin-resistant gene, a single copy of the *lacI* gene, and a single copy of the *argU* gene encoding the tRNA cognate to the rare AGA codon for Arg. As previously described, we scored the protein expression level from two transformants of the same plasmid on an integer scale from 0 (no expression) to 5 (highest expression), based on visual inspection of whole-cell lysates on Coomassie-blue-stained SDS-PAGE gels. There is an unmistakable difference between the 0 and 5 expression scores used for most of the analyses reported in this paper. A score of 5 indicates the target protein was the most abundant protein expressed in the cell, while a score of 0 indicates it was undetectable against the background of cellular proteins. The reproducibility of the integer scores in our large-scale data set was excellent, as analysed in detail previously³⁹. There was no difference between all measurements for over 70% of the genes and a maximum difference of one unit between all measurements for over 80% of the genes. When replicates gave different scores, the maximum score was used, because most sources of experimental error tend to reduce expression score, and bell-weather analyses reported in our previously published paper³⁹ showed a small increase in the significance of correlations when using maximum rather than mean score.

Computational modelling. Our binary multi-parameter logistic regression model gives θ , the logarithm of the ratio of the probabilities of obtaining the highest level of protein expression (P_{E5}) versus none (P_{E0}) from an mRNA sequence in the large-scale data set, as a linear function of generalized variables x_i :

$$\theta = \ln[P_{E5} / P_{E0}] = A + \sum_i \beta_i x_i$$

The probability of obtaining the highest level ($E = 5$) versus no ($E = 0$) protein expression from a given sequence is therefore given by:

$$\pi(\theta) = \frac{P_{E5}}{P_{E0} + P_{E5}} = \frac{\exp\{\theta\}}{1 + \exp\{\theta\}}$$

Note that, to capture nonlinear relationships between mRNA sequence parameters and outcome, the generalized variables x_i can represent mathematical functions of mRNA sequence parameters as well as those parameters themselves. We used the *R* statistics program⁵³ to compute the most probable values of the model parameters (A , β_i). Logistic-regression slopes $\beta_i > 0$ indicate that the probability of high expression increases as the associated variable increases in numerical value. (Note that, because ΔG increases in numerical value as folding stability decreases, a positive slope for free-energy terms indicates an increase in the probability of high expression as predicted folding stability decreases, while a negative

slope for these terms indicates an increase in the probability of high expression as predicted folding stability increases.) Our final model, which we call model M (Extended Data Table 1a and Fig. 4), is given in the main text, and the codon slopes β_c from this model are depicted in Fig. 3a. In principle, the probability of high protein expression can be increased by manipulating mRNA sequence properties to maximize the value of θ and thus π in the equations above using the parameters (A , β_i) from model M.

Inclusion of parameters was guided by the likelihood ratio test in conjunction with the AIC⁵⁴, a standard measure of whether an improvement in model quality exceeds that expected at random from increasing the number of degrees of freedom in the model. The likelihood ratio χ^2 (LR χ^2) is asymptotic to the χ^2 distribution and defined as the reduction in the deviance D of the observed data from the predictions of the model compared to the null model containing just the constant term A (in the first equation above), while the AIC is given by the LR χ^2 minus two times the number of degrees of freedom. The deviance is defined as:

$$D = -2 \sum_{j=1}^n [E_j \ln(\pi_j) + (1 - E_j) \ln(1 - \pi_j)]$$

This sum is conducted over the $n = 3,727$ proteins giving expression scores of 5 or 0 among the 6,348 in the large-scale protein expression data set, and the logistic variable E_j assumes values of 1 or 0 if protein '*j*' is expressed at the $E = 5$ or $E = 0$ levels, respectively. The variable $\pi_j = \pi(\theta_j)$ gives the predicted probability of obtaining expression of protein '*j*' at the $E = 5$ rather than $E = 0$ level according to the equations given above describing the multi-parameter binary logistic model. For the data set analysed in this paper, the deviance has values of 5,154 and 3,952 for the null model and our final model M, respectively (Extended Data Table 1a). In addition to using the AIC, we ensured that the final model is not over-fit via bootstrapping with replacement 1,000 times using the RMS package⁵⁵. This validation procedure is considered more robust than splitting the data set into training and test sets, which requires very careful selection of the test set.

The sequence parameters explored in the course of model development (Extended Data Table 1 and additional data not shown) included the length of the gene, the individual codon frequencies in-frame or out-of-frame in the entire gene, the individual codon frequencies in-frame calculated separately in the head and the tail or in the first and second halves of the coding sequence, di-codon frequencies, the statistical entropy of the codon sequence, the codon and amino acid repetition rates (defined below), the frequencies of the nucleotide bases at each codon position in the entire gene and in defined windows within its sequence, and a variety of predicted mRNA-folding energy parameters including those shown in Fig. 1 and Extended Data Fig. 2, which were evaluated individually and as statistical aggregates. The codon repetition rate r_{codon} and amino acid repetition rate r are defined as $< d_i^{-1} >$, where d_i is the distance at every position in the sequence to the next occurrence of the same species moving towards the 3' end of the gene. The value of d_i^{-1} is set to zero if the codon or amino acid does not occur again, so the value of r for the protein sequence LRPRL is the average of (1/4, 1/2, 0, 0, 0), which is 0.15. The sequence of the C-terminal LEHHHHHH affinity tag was omitted from all computational analyses to avoid biasing statistics on its constituent amino acids and codons. Because this sequence is present in every gene included in our large-scale protein expression data set, it cannot directly influence outcome on its own and can only have an influence via differential interaction with other sequence features. No evidence of such interactions was detected in bell-weather analyses including the tag sequence, so it was omitted in the final analyses reported in this paper.

The number of degrees of freedom for codon variables is one fewer than the number of non-stop codons because their frequencies f_c in a sequence must sum to 1 (that is, $\sum f_c = 1$). Therefore, for the analyses shown in Figs 3 and 4, we removed ATG, effectively constraining its slope to be zero (that is, $\beta_{\text{ATG}} = 0$) and its contribution to the model to be absorbed into the constant A . The inclusion of mean codon-slope variables s_{7-16} and s_{17-32} in model M uniformly reduces the individual codon slopes β_c to ~86% of their values when no mean-slope terms are included in the model, reflecting the disproportionate influence of codons near the 5' terminus compared to those in the rest of the gene (Extended Data Fig. 6). We tested expanded codons models including the next base or the previous base in addition to the in-frame codon, but these were rejected based on the AIC and bootstrap validation criteria described above.

We also examined introducing additional variables into model M (Extended Data Table 1b and additional data not shown). Adding the mean value of the predicted free energy of mRNA folding in the tail does not significantly improve the model, even though unstable folding in the tail correlates with reduced protein expression (Fig. 1g, h). Therefore, this correlation as well as those of the overall A, T, G and C content in the gene (Extended Data Fig. 2a–e) are captured more effectively by the cross-correlated sequence

parameters (Extended Data Figs 3 and 4) that are included in the model, suggesting that these other parameters are more influential mechanistically. Adding the mean slope of codons 2–6 does not produce a statistically significant improvement, and using this term instead of the base-composition terms in this region yields inferior results, consistent with the analyses shown in Extended Data Fig. 5. Finally, adding the frequency of the Shine–Dalgarno consensus AGGA in any frame (f_{AGGA} in Extended Data Fig. 2i, j and Extended Data Table 1b) fails to produce a statistically significant improvement. We also used the Bindigo program (<http://rna.williams.edu/>) to compute the binding energy of all hexamer sequences in a gene with the anti-Shine–Dalgarno sequence CACCUCCU, and neither the minimum nor the average value of the predicted free energy of hybridization to the anti-Shine–Dalgarno sequence has any correlation with protein expression level over our large-scale data set (Extended Data Table 1b).

Design of synonymous mRNA sequences. In the 6AA method, codons for six amino acids were changed to the single codon specified in Extended Data Table 2, which has a larger slope than that of any synonymous codon in our single-parameter binary logistic regression analyses (dark grey symbols in Fig. 3a). Although no explicit free energy optimization was performed with the 6AA method, it produced genes in which the predicted free energies of mRNA folding were more favourable than those in the naturally occurring starting sequences. In the 31C-FO method, predicted mRNA-folding energy was optimized while selecting codons from the 31 listed in Extended Data Table 2, which have slopes greater than zero in our single-parameter binary logistic regression analyses (dark grey symbols in Fig. 3a). The predicted free energy of folding of the head plus 5'-UTR (ΔG_{UH}) was maximized numerically (that is, to yield the least stable folding), while the predicted free energy of the folding in the tail was optimized to be near $-10 \text{ kcal mol}^{-1}$ in windows of 48 nucleotides. The 31C-FD used the same set of codons to produce genes in which the predicted free energy of folding was minimized numerically (that is, to yield the most stable folding).

Bacterial strains and growth media. The *E. coli* strain DH5 α was used for cloning. Expression experiments used *E. coli* strain BL21(DE3) pMGK (ref. 38). Ampicillin was added at $100 \mu\text{g ml}^{-1}$ for cultures harbouring pET21-based plasmids. Kanamycin was added at $25 \mu\text{g ml}^{-1}$ to maintain the pMGK plasmid. Bacterial growth for protein expression and northern blot experiments employing pET21-based plasmids was performed using the same medium and conditions that were used to generate our high-throughput protein-expression data set³⁸ (that is, MJ9 minimum medium⁵⁶ with 250 r.p.m. agitation at 37°C before induction at 17°C).

Plasmids. The pET-21 clones of the genes APE_0230.1 (*Aeropyrum pernix* K1), RSP_2139 from (*Rhodobacter sphaeroides*), SRU_1983 (*Salinibacter ruber*), SCO1897 (*Streptomyces coelicolor*) and *ycaQ* (*E. coli*) were obtained from the protein-production laboratory of the Northeast Structural Genomics Consortium (<http://www.NESG.org>) at Rutgers University (NESG targets Xr92, RhR13, SrR141, RR162 and ER449, respectively). The DNAs encoding the 6AA_T and 31C-FO_H/31C-FO_T variants of the genes were synthesized by GenScript. The head variants 31C-FO_H and 31C-FO_T were generated by PCR amplification using long forward primers containing an NcoI restriction site, the new head sequence, and a sequence complementary to the downstream region in the target gene. A plasmid containing the starting construct was used as DNA template for PCR amplification using the corresponding long forward primers and a reverse primer hybridizing at the 3' end of the target gene including the XhoI restriction site. The resulting PCR products were cloned using the In-Fusion kit (Clontech) into a pET-21 derivative linearized with NcoI and XhoI. The full protein-coding sequence in every plasmid was verified by DNA sequencing (Genewiz and Eton Bioscience) and corrected when necessary using the QuikChange II Site-Directed Mutagenesis kit (Agilent Technologies). The wild-type and 31C-FO_H/31C-FO_T (31C-FO_{H/T}) genes for SRU_1983, APE_0230.1 and *E. coli* YcaQ were re-cloned into a pBAD expression plasmid (Life Technologies) with a C-terminal hexa-histidine tag for transcription by the native *E. coli* RNA polymerase under control of an arabinose-inducible promoter; these experiments yielded similar results (Extended Data Fig. 6e, f) to those shown for the same genes under T7 polymerase control in a pET plasmid (Fig. 5 and Extended Data Fig. 6a–d). DNA sequences of the final constructs are provided in Supplementary Data File 3.

***E. coli* growth curves.** Overnight cell growth was measured by transferring 200 μl of each induced culture to a 96-well sterile plate (Greiner Bio-One) and covering each well with 50 μl of sterile paraffin oil. A negative control non-induced sample was loaded for each wild-type target. Duplicate wells were measured for each sample. Plates were loaded into a platereader (Biotek Synergy) at room temperature and shaken for 30 s. An initial $A_{600 \text{ nm}}$ reading was taken and then followed by 30 min of shaking until the next absorbance reading. Readings were repeated at 30 min intervals during 9 h of cell growth.

Analysis of protein expression *in vivo* from IPTG-inducible pET vectors. Starting cultures from a single colony were inoculated into 6 ml of LB media

containing $100 \mu\text{g ml}^{-1}$ of ampicillin and $30 \mu\text{g ml}^{-1}$ kanamycin. Cultures were grown at 37°C until highly turbid (4–6 h), then 40 μl was used to inoculate 2 ml of MJ9 chemically defined medium⁵⁶. This MJ9 pre-culture was grown overnight at 37°C. The next day, $A_{600 \text{ nm}}$ readings were taken of a 1:10 dilution of the turbid MJ9 pre-culture. This reading was used to calculate the volume of pre-culture necessary to normalize all cell samples to a starting culture density of 0.1 $A_{600 \text{ nm}}$ in 6 ml of fresh medium. The reinoculated culture was grown at 37°C until $A_{600 \text{ nm}}$ reached 0.5–0.7. Cells were then induced with 1 mM IPTG, with one duplicate tube for each wild-type gene not induced to serve as a negative control. After induction, 200 $\mu\text{l} \times 2$ of each culture was removed and placed into a sterile 96-well plate to monitor cell growth rate (see above). The remaining 5.6 ml of induced samples were then transferred to 17°C and shaken overnight. The next day, samples were removed from the shaker, placed on ice, and final $A_{600 \text{ nm}}$ was measured. Cells were centrifuged in 14-ml round-bottom Falcon tubes at 5,300g for 10 min, and the pellets were resuspended in 1.2 ml of lysis buffer (30 mM NaCl, 10 mM 2-mercaptoethanol, 50 mM NaH₂PO₄, pH 8.0) and then transferred to 1.5 ml Eppendorf tubes on ice. Lysis was accomplished by sonication on ice, using a 40 V setting ($\sim 12 \text{ W pulse}$) and pulsing for 1 s followed by a 2 s rest, for a total of 40 pulses. Then 120 μl of each lysed culture was mixed with 40 μl of 4 \times Laemmli buffer, and samples were analyzed using SDS–PAGE (Bio-Rad, Ready Gel, 15% Tris–HCl), with Bio-Rad Precision Plus All Blue Standard markers. Final $A_{600 \text{ nm}}$ measurements were used to calculate the load volume for each individual sample, normalizing all samples to the density of the least turbid of each unique target. We verified the integrity of the plasmids after growth and induction by DNA sequencing (Genewiz and Eton Bioscience). Every result was confirmed by repeating the experiment.

Analysis of protein expression *in vivo* from arabinose-inducible pBAD vectors. Conducting experiments at physiological protein expression levels (Extended Data Fig. 6e, f) required considerable changes in methods compared to the experiments conducted in pET vectors that were used to generate our large-scale protein-expression data set and the data shown in Fig. 5 and Extended Data Figs 6a, b and 7. Because mRNA expression from IPTG-controlled promoters tends to occur in an all-or-none fashion^{60,61}, it is not practical to control the level of mRNA expressed from pET vectors. Therefore, we re-cloned three pairs of synonymous native and codon-optimized 31C-FO_{H/T} genes with C-terminal hexahistidine tags under control of the arabinose-inducible promoter in a pBAD vector⁶², which provides a more gradual increase in expression as arabinose concentration is raised. This promoter drives transcription using the endogenous *E. coli* RNA polymerase rather than T7 RNA polymerase, which is employed by the pET vectors used for all other expression experiments reported in this paper. Because transcription from the arabinose promoter is repressed by glucose, which is the carbon source in the chemically defined MJ9 medium used for our pET experiments, we instead used LB as the growth medium for pBAD experiments, which were conducted in BL21 pMGK cells (that is, an isogenic *E. coli* strain except for the removal of the λ (DE3) prophage carrying the gene for T7 RNA polymerase). Furthermore, because the arabinose inducer can be depleted during long growth periods, we evaluated expression after relatively short 1–4 h induction times during log-phase growth rather than after overnight growth into stationary phase, which was used for our pET experiments. We also changed the growth temperature during induction from 17°C for pET experiments to 37°C for pBAD experiments. Non-induced controls were grown in medium containing 0.4% glucose (+Glc). When the $A_{600 \text{ nm}}$ of the cultures reached 0.6, transcription of the target genes was induced for 1 h using final arabinose concentrations of 0.001% (w/v) for APE_0230.1 and 0.01% (w/v) for SRU_1983 and *E. coli* YcaQ (+Ara).

***In vitro* transcription.** The pET21 plasmids containing optimized or unoptimized inserts were digested with BspI, phenol–chloroform purified, and concentrated by ethanol precipitation. From the digested samples, 2 μg was added to the RiboMax kit (Promega), and *in vitro* transcription with bacteriophage T7 RNA polymerase was conducted according to the manufacturer's protocol. Upon completion of the reaction, samples were treated with DNase (Promega), isopropanol precipitated, and resuspended in RNA Storage Solution (Ambion). Transcript size and purity were verified by agarose gel electrophoresis with ethidium bromide staining. For kinetic analyses, 20- μl reactions with T7 polymerase were assembled and started by addition of 1 μg of template DNA. A 4.5- μl sample of each reaction was removed at 0-, 5-, 10- and 30-min time points for analysis on denaturing formaldehyde-agarose gels. Each experiment was conducted at least twice.

***In vitro* translation.** *In vitro* translation assays of the purified mRNAs were performed with the PURExpress system (New England Biolabs) using L-[³⁵S]methionine premium (PerkinElmer). Each 25- μl reaction contained 10 μl of solution A, 7.5 μl of solution B and 2 μl of [³⁵S]methionine (10 μCi). The reactions were started by adding 2 μl of purified mRNA (4 $\mu\text{g} \mu\text{l}^{-1}$) and incubating at 37°C. Aliquots of 5 μl were withdrawn from the reactions at 15, 30, 60 and 90 min, and translation was stopped by adding 10 μl of 2 \times Laemmli and heating for 2 min at 60°C. Then 14 μl

of each aliquot was run on a 4–20% SDS–PAGE gel (Bio-Rad) with Bio-Rad Precision Plus All Blue Standard markers. The gel was dried on Whatman filter paper and subjected to autoradiography. Each reaction was repeated at least twice.

Northern blot analyses. The probe was designed as the reverse complement of the 71-nucleotides of the 5′-UTR of the pET21 vector, and it was synthesized by Eurofins. The probe was labelled with biotin using the BrightStar Psoralen-Biotin Nonisotopic Labelling Kit. BL21(DE3) pMGK *E. coli* containing the plasmid of interest were grown overnight in LB at 37 °C with shaking. Cultures were diluted 1:50 into MJ9 media and grown overnight at 37 °C with shaking. The next day, the cultures were diluted to an $A_{600\text{ nm}}$ of 0.15 in MJ9 media and allowed to grow to an $A_{600\text{ nm}}$ of 0.6–0.7 before induction with 1 mM IPTG. Samples were taken at the indicated time points and RNAs were stabilized in two volumes of RNAProtect Bacteria Reagent. After pelleting, samples were lysozyme digested (15 mg ml^{-1}) for 15 min, and RNAs were purified using the Direct-zol RNA Miniprep Kit and TRI-Reagent. Approximately 1–2 µg of total RNA per sample was separated on a 1.2% formaldehyde-agarose gel in MOPS-formaldehyde buffer. RNA integrity was verified by ethidium bromide staining. RNA was then transferred to a positively charged nylon membrane using downward capillary transfer with an alkaline transfer buffer (1 M NaCl, 10 mM NaOH, pH 9) for 2 h at room temperature. RNAs were crosslinked to the membrane using 1,200 µJ ultraviolet irradiation (Stratalinker). Membranes were pre-hybridized in Ultrahyb hybridization buffer for 1 h at 42 °C in a hybridization oven. Heat-denatured, biotin-labelled probe was then added to 10–20 pM final concentration and hybridized overnight at 42 °C. Membranes were washed twice in buffer ($0.2\times$ SSC, 0.5% SDS), and probe signal was detected using the BrightStar BioDetect kit, as per protocol, via exposure to film. Each northern blot experiment was repeated at least twice.

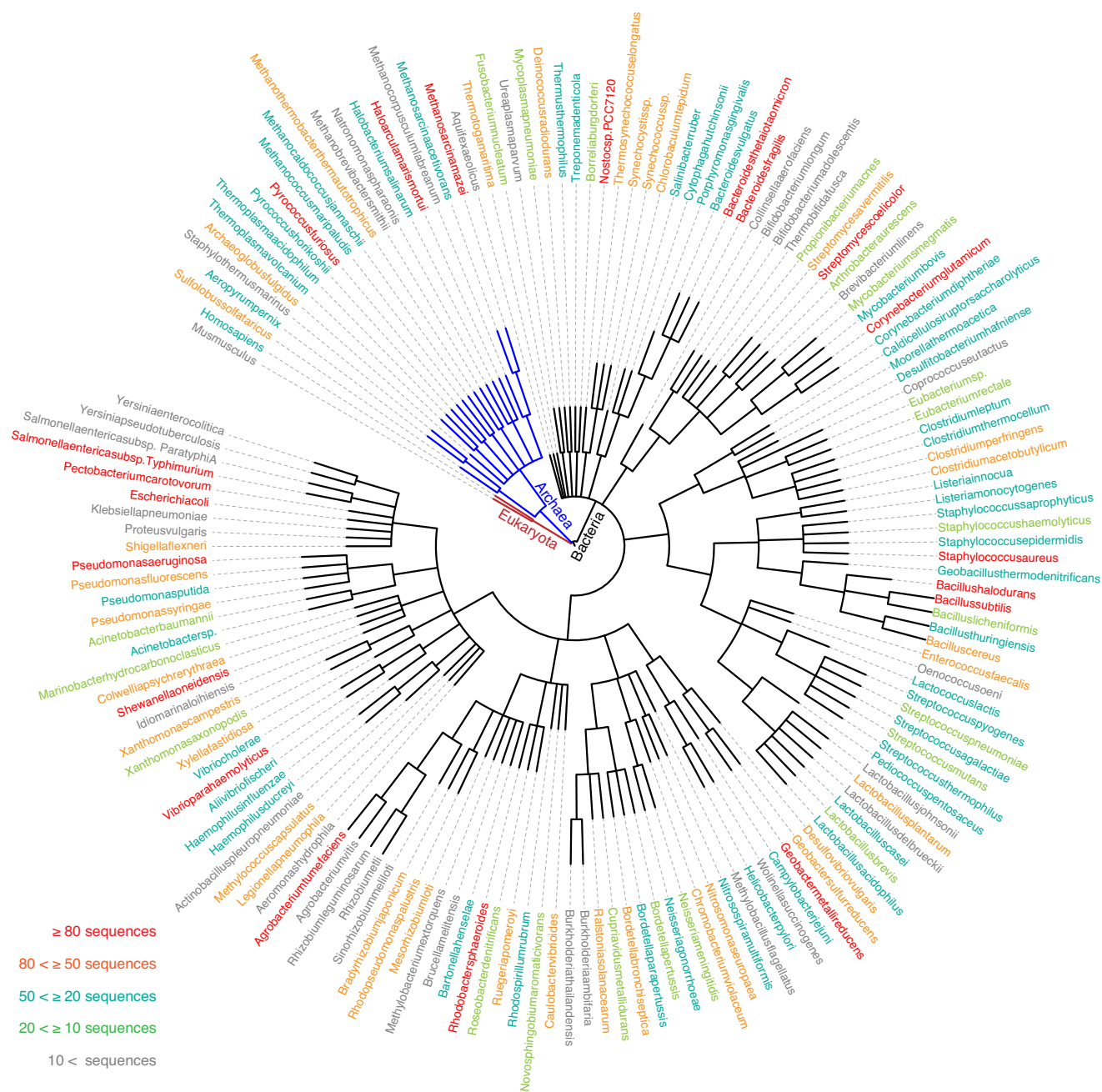
RNA extraction and microarray analyses. *E. coli* MG1655 cells were cultured in M9 0.4% glucose minimum media to a final $A_{600\text{ nm}}$ of 1.0. Cells were treated with RNA Protect Bacteria Reagent (Qiagen), and RNA extracted using the RNeasy Mini Kit (Qiagen) was reverse-transcribed using SuperScript II Reverse Transcriptase (Invitrogen) followed by treatment with RNaseH (Invitrogen) and RNaseA (EpiCentre). The resulting cDNA preparation was purified using the MinElute Purification Kit (Qiagen) and then fragmented into 50–200-bp fragments using DNaseI (EpiCentre). Biotinylation was performed with Terminal Deoxynucleotidyl Transferase (New England Biolabs) and Biotin- N^6 -ddATP (Enzo Life Sciences). Biotinylated cDNA was hybridized on Affymetrix *E. coli* 2.0 arrays by the Gene Expression Center at the University of Wisconsin Biotechnology Center. Raw data (.cel) files were analysed using the RMA (Robust Multi-chip Average) algorithm in the Affymetrix Expression Console.

Classification of cytoplasmic proteins in *E. coli* MG1655. All predicted proteins in the version of the genome in the Ecocyc database⁵⁷ were analysed using the

programs LipoP⁵⁸ and TMHMM⁵⁹, and those without a predicted transmembrane helix or a predicted signal peptide were classified as cytoplasmic proteins and included in the analyses in Fig. 6.

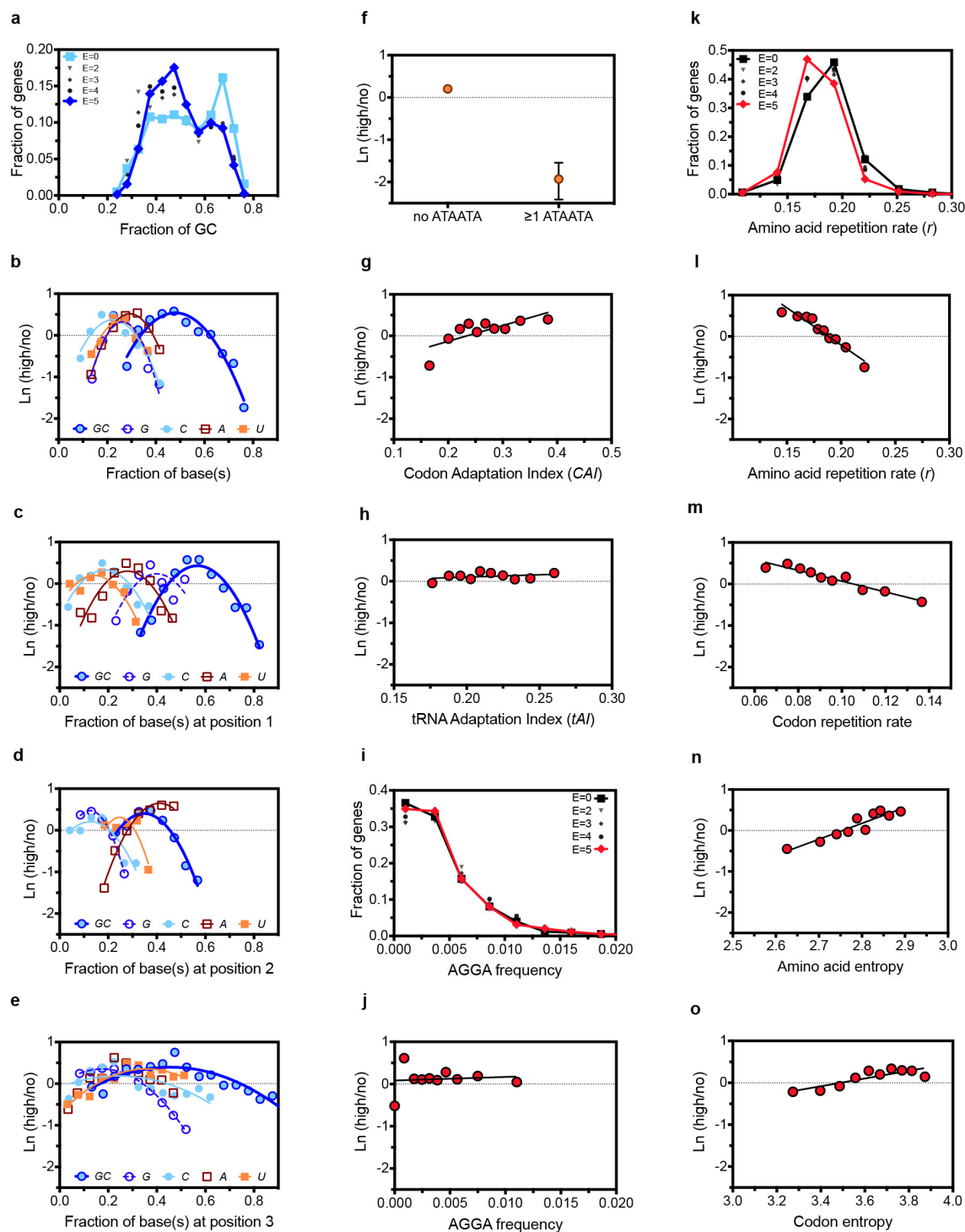
Analysis of mRNA lifetime data sets. We analysed the data sets published previously⁴⁴ in which RNA-seq was used to quantify global mRNA levels as a function of time after treatment of either exponential or early stationary phase cultures with the transcription-initiation inhibitor rifampicin. To avoid potential complications arising from the encoding of multiple proteins in polycistronic transcripts, we limited our analyses to monocistronic transcripts, which constituted 76% and 82% of the mRNAs for which lifetimes were measured in exponential and stationary phase, respectively. The analyses presented in Fig. 6c, d were also limited to predicted cytoplasmic proteins to avoid possible biases from systematically lower expression of integral membrane proteins and secreted proteins. The set of genes for which Chen *et al.*⁴⁴ were able to measure lifetime is strongly biased towards more abundant mRNAs, and the measured lifetimes in both the exponential and stationary phase data sets are also strongly correlated with steady-state concentrations (data not shown).

51. Xiao, R. *et al.* The high-throughput protein sample production platform of the Northeast Structural Genomics Consortium. *J. Struct. Biol.* **172**, 21–33 (2010).
52. Acton, T. B. *et al.* Preparation of protein samples for NMR structure, function, and small-molecule screening studies. *Methods Enzymol.* **493**, 21–60 (2011).
53. R Development Core Team. *A Language and Environment for Statistical Computing*; <http://www.r-project.org/> (2012).
54. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Auto. Con.* **19**, 716–723 (1974).
55. Harrell, F. E. Jr. *R package version 4.2-0*; <http://CRAN.R-project.org/package=rms> (2014).
56. Jansson, M. *et al.* High-level production of uniformly ^{15}N - and ^{13}C -enriched fusion proteins in *Escherichia coli*. *J. Biomol. NMR* **7**, 131–141 (1996).
57. Keseler, I. M. *et al.* EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* **41**, D605–D612 (2013).
58. Juncker, A. S. *et al.* Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* **12**, 1652–1662 (2003).
59. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
60. Novick, A. & Weiner, M. Enzyme induction as an all-or-none phenomenon. *Proc. Natl Acad. Sci. USA* **43**, 553–566 (1957).
61. Jensen, P. R., Westerhoff, H. V. & Michelsen, O. The use of lac-type promoters in control analysis. *Eur. J. Biochem.* **211**, 181–191 (1993).
62. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).



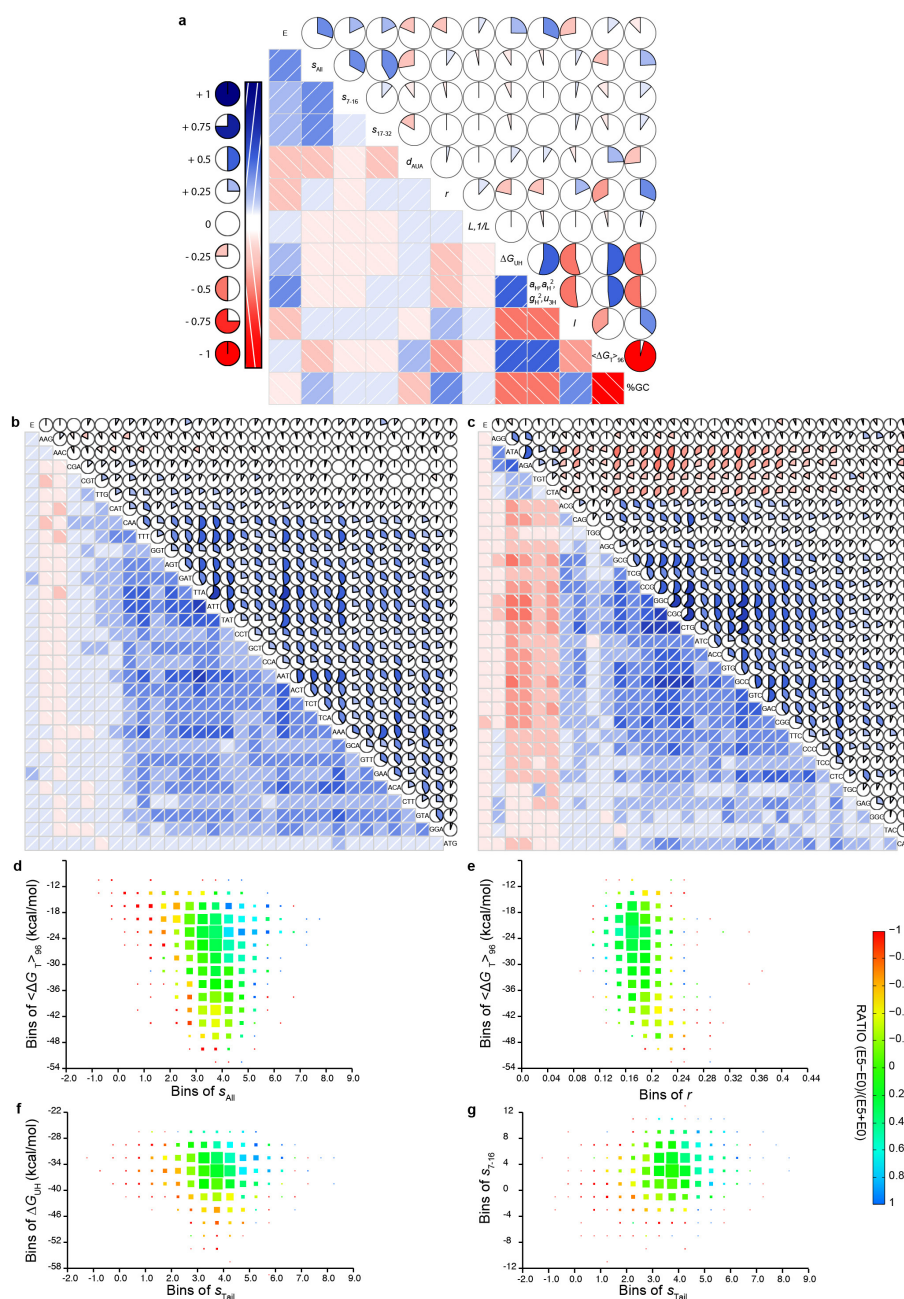
Extended Data Figure 1 | Phylogenetic distribution of the proteins in the large-scale protein expression data set. The colours in the cladogram encode the number of genes/proteins from each organism, as indicated by the legend. The data set includes 47 from eukaryotes (45 from humans

and 2 from mouse), 809 from archaeobacteria, and 96 from *E. coli*, with the remainder coming from other eubacteria. The organism contributing the largest number of proteins to the data set is the eubacterium *Bacteroides thetaiotaomicron* (150 proteins).



Extended Data Figure 2 | Relationships between additional mRNA sequence parameters and results in the large-scale protein expression data set. **a, i, k,** Histograms showing for each expression score the distribution of the overall G+C frequency (**a**), the frequency in all reading frames of the AGGA core sequence of the Shine–Dalgarno ribosome-binding sequence (**i**), and the amino acid repetition rate r (**k**; see Methods for definition). The parameter distributions in the $E = 5$ and $E = 0$ categories ($n = 3,727$ for both combined) are shown in **a** in dark and light blue, respectively, and in **i** and **k** in red and black, respectively. The symbols used for the histograms for the intermediate expression scores ($n = 2,621$ for all combined) are indicated in the legend for each panel. **b–h, j, l–o,** Plots showing the logarithm of the ratio of the number of proteins with $E = 5$ versus $E = 0$ scores as a function of parameter value. **b,** Data for the overall frequencies of the four individual nucleotide

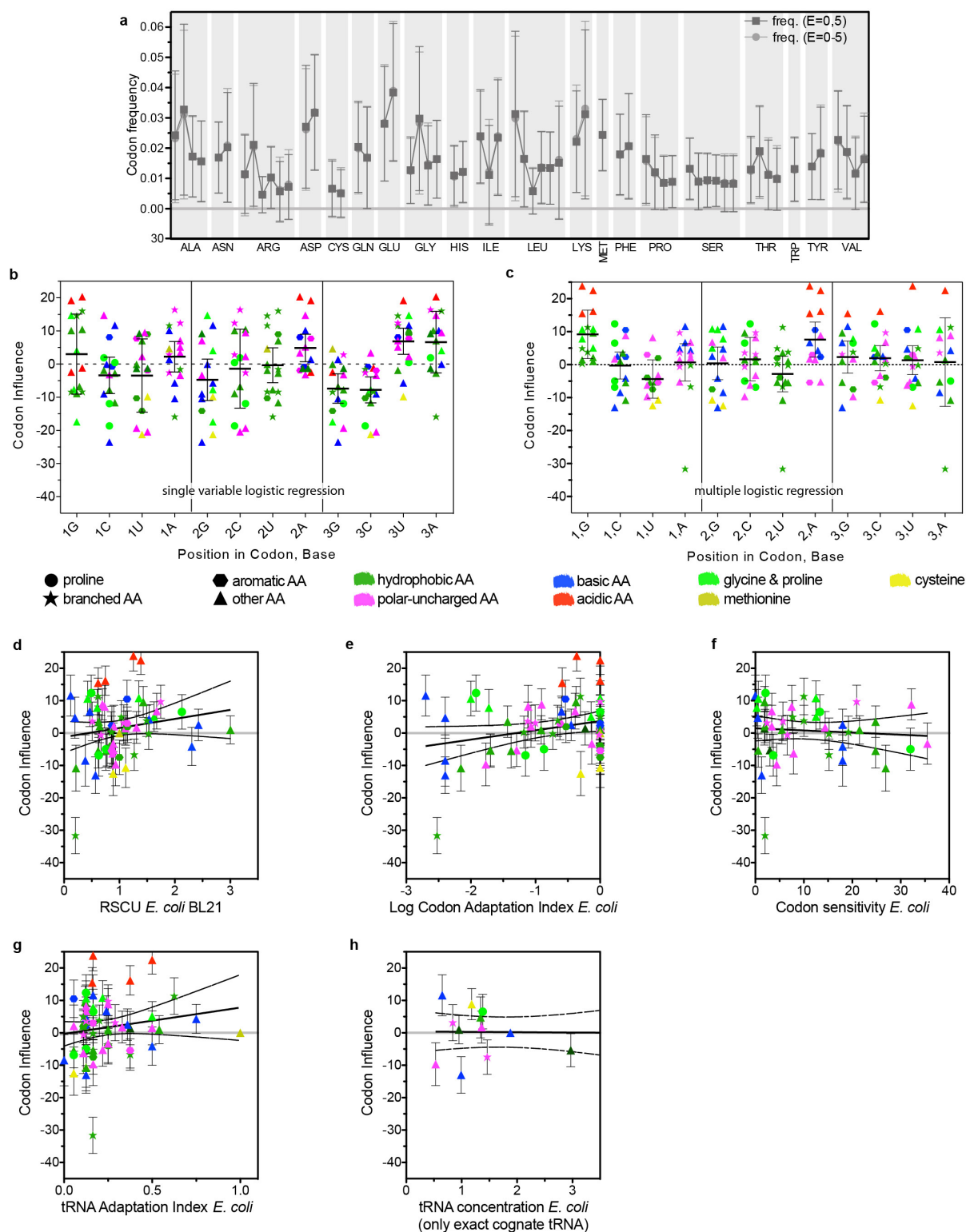
bases as well as the combined G + C frequency (labelled GC). **c–e,** The equivalent data separately for the first (**c**), second (**d**) and third (**e**) positions in the codons in the genes. **f,** Data for genes either not containing or containing at least one occurrence of the ATA–ATA di-codon ($P = 2 \times 10^{-32}$). The error bars in this panel represent 95% confidence limits calculated from bootstrapping; the error bars for the genes without any occurrence of this di-codon are smaller than the size of the symbol. **g, h,** Data for the codon adaptation index¹⁴ (**g**) and tRNA adaptation index¹⁶ (**h**). **j,** Data for the frequency in all reading frames of the sequence AGGA. **l, m,** Data for the amino acid repetition rate r (**l**) and the codon repetition rate (**m**). **n, o,** Data for the statistical entropy of the amino acid (**n**) and codon sequences (**o**). The data in **a–e, i** and **k** are binned in equal ranges of the parameter value, while the data in **g, h, j** and **l–o** are binned in deciles containing equal populations.



Extended Data Figure 3 | Correlations between sequence parameters in the genes included in the large-scale protein expression data set.

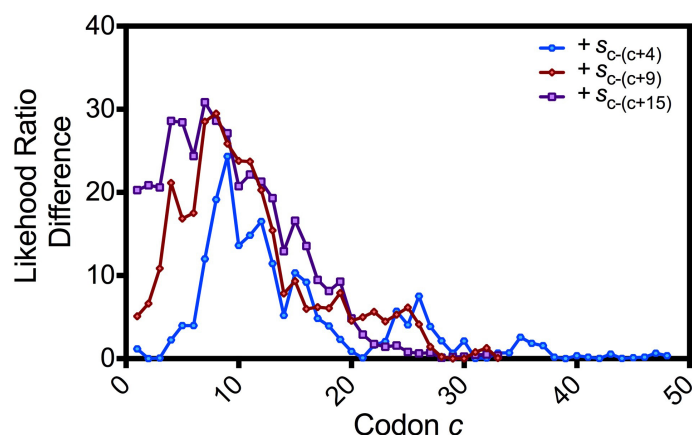
a–c, Corrgrams representing the signed Pearson correlation coefficients between different mRNA sequence parameters in the genes in the $E = 0$ plus $E = 5$ categories in the data set ($n = 3,727$ for the two combined). The colour-coding is defined schematically on the left in **a**, with blue being used for positively correlated variables, red for negatively correlated variables, and white for uncorrelated variables. In **a**, E represents the expression score in the binary categories (0, 5), s_{all} represents the mean value of our new codon-influence metric (coloured symbols in Fig. 3a) over the entire gene (without the LEHHHHH tag), s_{7-16} and s_{17-32} represent the mean values of this metric for codons 7–16 and 17–32, respectively, ΔG_{UH} represents the predicted free energy of mRNA folding for the 5'-UTR from the pET21 expression vector plus the first 48 nucleotides in the gene, $<\Delta G_T>_{96}$ represents the mean value in the remainder of the gene of the predicted free energy of folding in 50% overlapping windows of 96 nucleotides, I represents an indicator variable that assumes a value of 0 or 1 if ($\Delta G_{UH} < -39 \text{ kcal mol}^{-1}$) and ($\%GC_{2-6} > 0.65$), d_{AUA} assumes a value of 0 or 1 if there is at least one occurrence of the ATA–ATA di-codon, r represents the codon repetition rate (see Methods), and %GC represents the percentage content of G plus C bases in the gene. The variables a_H , a_H^2 , g_H^2 and u_H represent

monomial functions of the fractional content of A, G and U bases in codons 2–6; the correlation coefficient for these nucleotide-composition terms was calculated using their sum weighted by their optimized coefficients from model M (Fig. 4 and Extended Data Table 1a), as given in the equation in the main text. **b**, Data for the frequencies of the codons positively correlated with expression score E . **c**, Data for the frequencies of the codons negatively correlated with expression score E . **d–g**, Two-dimensional histograms illustrating the dependence of results in the large-scale protein-expression data set on pairs of sequence parameters. The colours encode the fractional excess of proteins with $E = 5$ versus $E = 0$ scores (that is, $(\#E5 - \#E0)/(\#E5 + \#E0)$), as calibrated by the scale bar on the right. The area of each square is proportional to the number of proteins in that bin in the two-dimensional parameter space. The variables s_{all} , s_{7-16} and s_{tail} represent, respectively, the mean values of our new codon-influence metric for the entire gene, for codons 7–16, and for all of the remaining codons downstream in the gene. ΔG_{UH} represents the predicted free energy of mRNA folding for the 5'-UTR from the pET21 expression vector plus the first 48 nucleotides in the gene, $<\Delta G_T>_{96}$ represents the mean value in the remainder of the gene of the predicted free energy of folding in 50% overlapping windows of 96 nucleotides, and r represents the amino acid repetition rate (as defined in Methods).



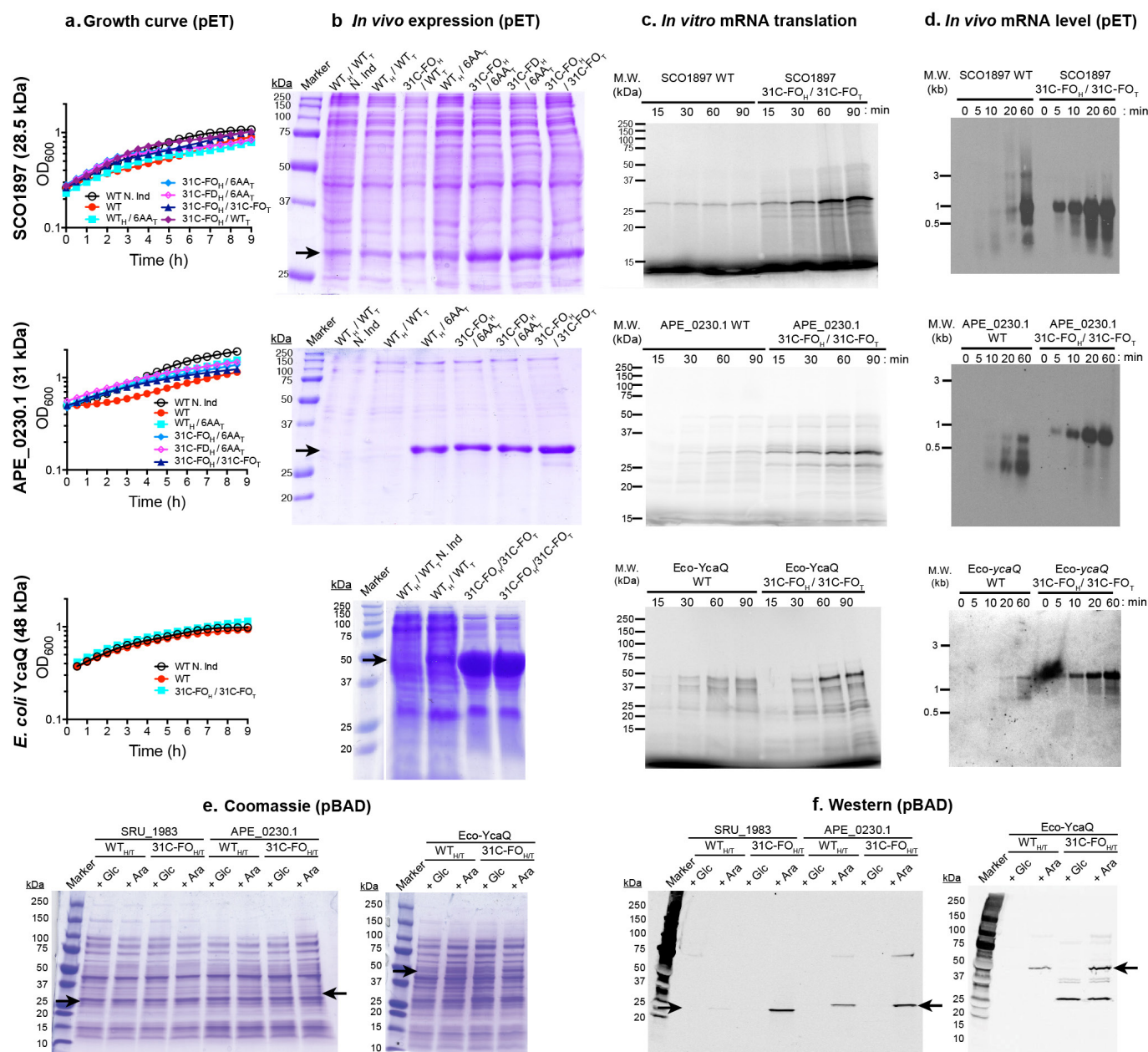
Extended Data Figure 4 | Relationship of the new codon-influence metric to parameters assumed to influence translation efficiency in previous literature. **a**, Average frequency of each non-stop codon in the genes in just the $E=0$ plus $E=5$ categories (dark grey) or in the $E=0$ through $E=5$ categories (light grey), with error bars representing the s.d. of the frequency among the genes in each set. **b**, Codon slopes from single-variable binary logistic regressions (dark grey symbols in Fig. 3a) segregated according to the identity of the nucleotide at each of the three positions in the codon. These slopes come from single-variable linear logistic regressions that were performed separately for each of the individual 61 non-stop codons. **c**, Codon slopes from the simultaneous

multi-parameter binary logistic regression model M (Extended Data Table 1a and coloured symbols in Fig. 3a) segregated according to the identity of the nucleotide at each of the three positions in the codon. **d–h**, The codon slopes from model M plotted versus the relative synonymous codon usage (RSCU) in *E. coli* BL21 (**e**), the codon adaptation index¹⁴ in *E. coli* K12 (**f**), the codon sensitivity²⁴ in *E. coli* K12 (**d**), the tRNA adaptation index¹⁶ in *E. coli* K12 (**g**), and the concentration of exactly cognate tRNAs²³ in *E. coli* K12 (**h**). The shapes and colour-coding of the symbols in **b–h**, which are the same as in Fig. 3, encode structural and qualitative chemical characteristics of the amino acids.



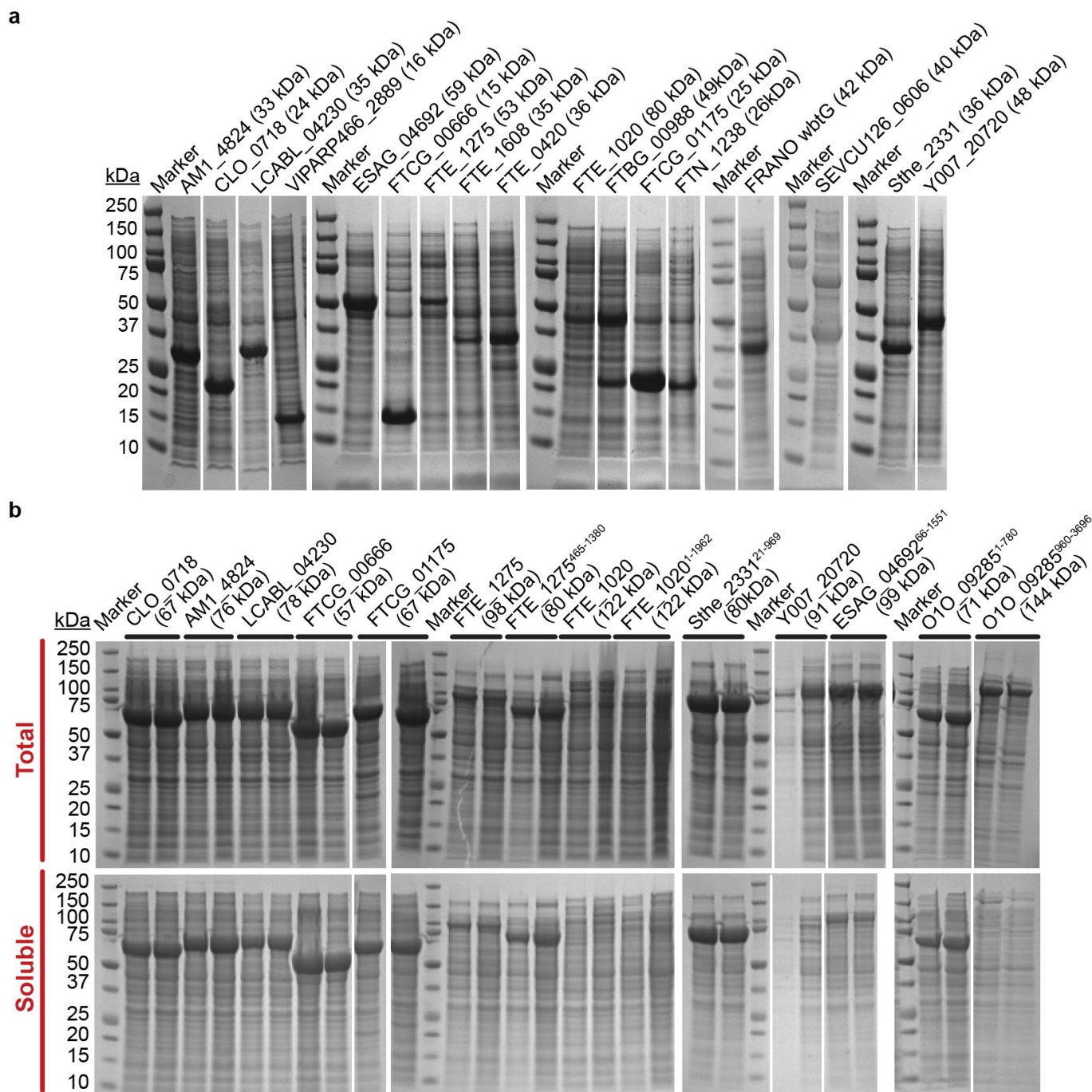
Extended Data Figure 5 | Variation in codon influence as a function of position in the coding sequence. Plots showing the reduction in

the deviance of the computational model resulting from adding a term representing the average value of the codon slope (coloured symbols in Fig. 3a) in a window 5, 10 or 16 codons wide starting at the position indicated on the abscissa (that is, c through $(c+4)$ in blue, c through $(c+9)$ in red, or c through $(c+15)$ in purple, respectively, with c representing the number of the first codon in the window). The reduction in deviance was calculated relative to a base model containing codon frequencies in the entire coding sequence, head nucleotide composition terms (a_H , a_H^2 , u_{3H} and g_H^2), the predicted free energy of RNA folding in the head plus the 5'-UTR (ΔG_{UH}), the binary indicator variable for head folding effects I , the binary variable indicating the occurrence of an AUAUA di-codon d_{AUA} , and the codon repetition rate r ($n = 3,727$). The mean slope of codons 2–6 presumably does not improve the model because the head-composition terms rather than codon content dominate the influence of this region on protein-expression level. This effect also probably accounts for the peaks in the $s_{c-(c+9)}$ and $s_{c-(c+15)}$ plots for windows starting at codon 7. For reference, adding s_{7-16} and s_{16-32} terms to model M contributes 29.7 points ($P = 5 \times 10^{-8}$) and 12 points ($P = 5 \times 10^{-4}$) of model deviance, respectively (Extended Data Table 1 and Fig. 4a). Dropping out terms to measure their influence (Fig. 4a) shows every codon contributes on average $(423.7/270) = 1.6$ deviance units, while codons 7–16 each contribute on average an additional $(29.6/10) = 3.0$ deviance units. Therefore, individual codons at positions 7–16 are approximately three times more influential than those in the tail of the gene.



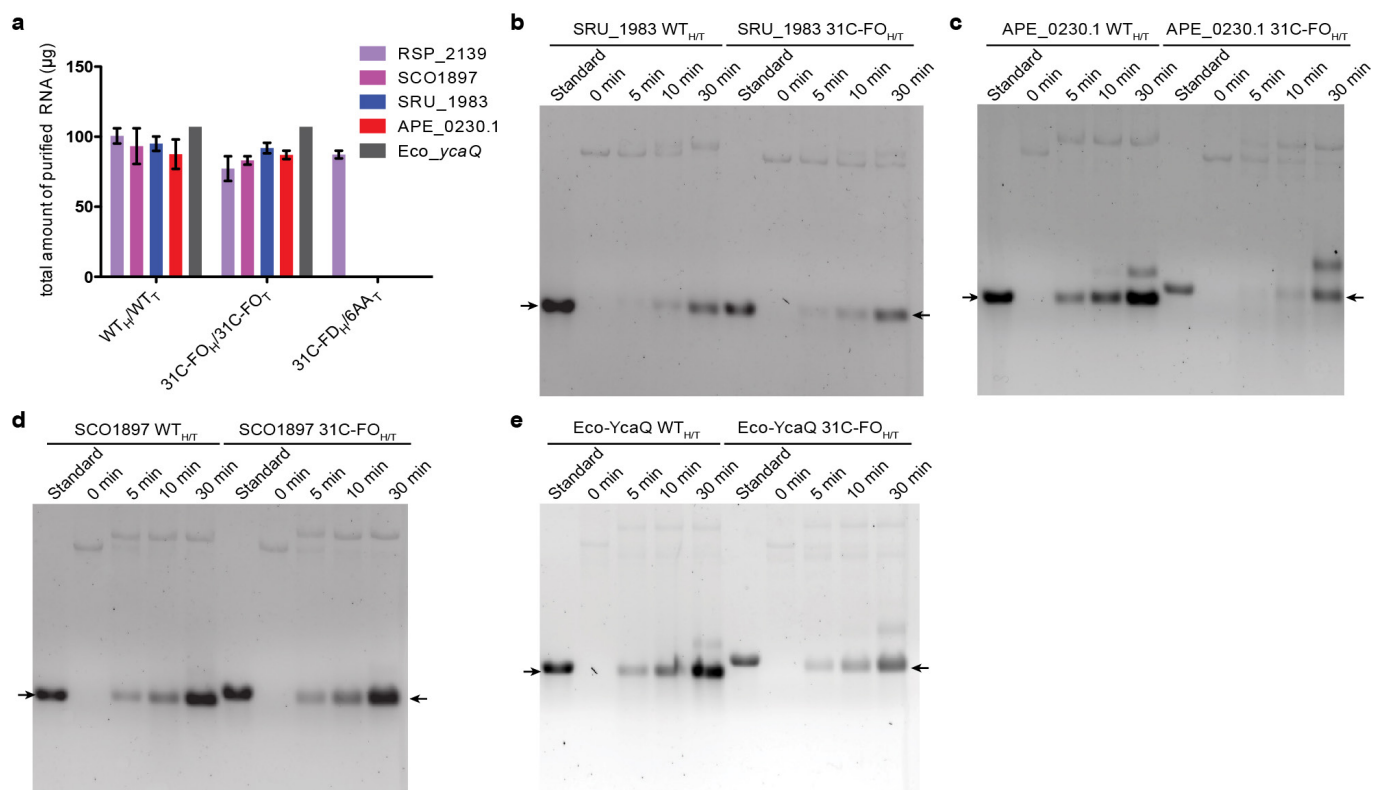
Extended Data Figure 6 | Further experiments on synthetic genes designed to enhance protein expression. a–d, Data for three additional proteins equivalent to the data presented in Fig. 5. The *in vivo* and *in vitro* expression properties from pET vectors are compared for inefficiently translated native (WT) genes and synonymous genes redesigned in the head or the tail or both using the 6AA, 31C-FO or 31C-FD methods. The type of sequence in the head (H) is indicated separately from that in the tail (T), and the name of the target protein is indicated on the left on each row. **a**, *E. coli* BL21(DE3) host cell growth curves at room temperature after induction of the target gene at time zero in chemically defined MJ9 medium. **b**, Coomassie-blue-stained SDS–PAGE gels of whole cells after overnight induction at 17°C, with the amount loaded in each lane normalized to the A_{600} nm of the culture at the time of harvest. Black arrows indicate the migration positions of the target proteins. **c**, Autoradiographs of SDS–PAGE gels of *in vitro* translation reactions using fully purified translation components in the presence of [35 S]methionine. Each reaction contained an equal amount of purified mRNA that was transcribed *in vitro* using T7 RNA polymerase. **d**, Northern blot analyses of the mRNA for the target protein after induction of expression *in vivo*. An equal amount of total RNA was loaded in each lane, and blots were hybridized with a probe matching the 5′-UTR. **e**, **f**, Coomassie blue stained SDS–PAGE gels (**e**) and anti-tetrahistidine western blots

(**f**) showing that gene optimization has equivalent effects at physiological protein expression levels. Pairs of synonymous native (WT) and codon-optimized 31C-FO_{H/T} genes with C-terminal hexahistidine tags were re-cloned under control of the arabinose-inducible promoter in a pBAD vector⁶², and the concentration of arabinose in the growth medium was adjusted so the 31C-FO_{H/T} genes yielded protein expression in the physiological range as assessed from Coomassie blue stained SDS–PAGE gels of whole cell extracts. Black arrows indicate locations of the induced target proteins. Substantially lower protein expression from the wild-type genes compared to the synonymous 31C-FO_{H/T} genes in these experiments demonstrates that equivalent codon-usage effects are observed when proteins are overexpressed using a pET vector or expressed at roughly physiological levels using a pBAD vector, despite changes explained in the online Methods in the polymerase used to transcribe the genes, the medium used to grow the cells, and the timescale and temperature of the protein-induction process. The constitutively expressed ~25-kDa protein that reacts with the anti-tetrahistidine antibody in the cells containing the 31C-FO_{H/T} gene for YcaQ is probably an amino-terminally truncated protein synthesized from a 5′-truncated mRNA transcribed from an internal promoter sequence fortuitously introduced into this synthetic gene. Uncropped scans of the gels shown here are included in Supplementary Fig. 1.



VIPARP466_2889 from *Vibrio parahaemolyticus*; AM1_4824 from *Acaryochloris marina* MBIC11017; CLO_0718 from *Clostridium botulinum* E1; ESAG_04692 from *Escherichia* sp. 3_2_53FAA; FTECG_00666 and FTECG_01175 from *Francisella tularensis* subsp. novicida GA99-3549; FTE_1275, FTE_1608, FTE_0420 and FTE_1020 from *Francisella tularensis* subsp. novicida FTE; FRANO_wbtG and A1DS62_FRANO from *Francisella novicida*; FTBG_00988 and A7JEH2_FRATL from *Francisella tularensis* subsp. tularensis FSC033; FTN_1238 from *Francisella tularensis* subsp. novicida U112; O1O_09285 from *Pseudomonas aeruginosa* MPAO1/P1; Sthe_2331 from *Sphaerobacter thermophilus* DSM20745/S6022; SEVCU126_0606 from *Staphylococcus epidermidis* VCU126; and Y007_20720 from *Salmonella enterica* subsp. enterica serovar Montevideo 507440-20.

LCABL_04230 from *Lactobacillus casei* BL23;



Extended Data Figure 8 | Yield of mRNA from *in vitro* transcription using purified T7 RNA polymerase. **a**, Final yield of mRNA purified from reactions conducted under identical conditions, as described in the Methods. The yields were calculated from the optical density at 260 nm. **b–e**, Kinetic analyses of *in vitro* transcription reactions using formaldehyde-agarose gel electrophoresis. Samples were taken at

0, 5, 10 and 30 min. The gels were stained with ethidium bromide. The 'standard' lane contains 1 μg of the same mRNA after purification to enable calibration for differences in the sensitivity of the molecules to staining. Reactions were started by addition of the wild-type or 31C-FO_H/31C-FO_T (31C-FO_{H/T}) linearized plasmids encoding SRU_1983 (**b**), APE_0230.1 (**c**), SCO1897 (**d**), or Eco-YcaQ (**e**).

Extended Data Table 1 | Development and analysis of the simultaneous multi-parameter binary logistic regression model

a

Abr.	Model	LR χ^2	d.f.	Δ AIC
G	ΔG_{UH}	241.6	1	239.6
I	$(\%GC_{2-6} > 0.62)(\Delta G_{UH} < -39 \text{ kcal/mol})$	313.0	1	311.0
H	$head_{2-6} = a_H + a_H^2 + u_{3H} + g_H^2$	378.7	4	370.7
	H+I	472.0	5	462.0
	G+I+H	483.1	6	471.1
	Head=G+I+H+ s_{7-16}	620.0	7	606.0
ℓ	$L+L^{-1}$	31.6	2	27.6
	entropy of codons	25.9	1	23.9
	r_{codon}	66.7	1	64.7
	entropy of amino acids	74.1	1	72.1
	$r_{\text{codon, Half 1}} + r_{\text{codon, Half 2}}$	68.5	2	64.5
	$r_{AA, \text{Half 1}} + r_{AA, \text{Half 2}}$	117.6	2	113.6
r	r_{AA}	122.3	1	120.3
d	$d_{AUA} = (\#ATAATA > 0)$	140.6	1	138.6
	$\langle \Delta G_T \rangle_{96} + \langle \Delta G_T \rangle_{96}^{-1}$	157.9	2	153.9
	Amino acids	400.3	19	362.3
	$\sum_{p=1,2,3} (A_p + A_p^2 + G_p + G_p^2 + U_p + U_p^2)$	446.9	18	410.9
	codons + following base (CCCB)	859.2	243	371.2
	codons + preceding (BCCC)	872.3	243	386.3
	codons frame+1	506.1	63	380.1
	codons frame+2	506.2	63	380.2
	codons, halves 1 and 2	757.6	120	517.6
C	codons in frame	637.5	60	517.5
	Tail=C+r+d	759.2	62	635.2
	Tail=C+r+d+ s_{17-32}	769.1	63	643.1
	$C_{\text{Half 1}} C_{\text{Half 2}} G \ell r d$	1083.9	125	833.9
	$C_{1-16} C_{17-} G \ell r d$	1116.2	125	866.2
	$C_{1-6} C_{7-} G \ell r d$	1203.9	125	953.9
	CGIH $\ell r d$	1194.5	70	1054.5
	CGIH $\ell r d$ + s_{7-16}	1224.2	71	1082.2
M	CGIH $\ell r d$ + $s_{7-16} + s_{17-32}$	1236.2	72	1092.2

b

Model	LR χ^2	p	Model	LR χ^2	p
M+SD _{ave}	+0.02	0.90	M+ s_{2-6}	+0.30	0.58
M+SD _{max}	+0.47	0.74	M+ r_{codon}	+0.14	0.74
M+ f_{AGGA}	+1.27	0.26	M+%GC	+2.14	0.14
M+CAI	+0.04	0.85	M+ g_{3T}	+2.40	0.12
M+CAI ₁₋₁₆	+0.58	0.45	M+ a_T	+4.10	0.04
M+tAI	+1.53	0.22	M+ $\langle \Delta G_T \rangle_{96}$	+1.39	0.24
M+tAI ₁₋₁₆	+2.44	0.12	M+ $\langle \Delta G_T \rangle_{96}^{-1}$	+5.92	0.01

a, Summary of model development calculations. The symbols defined in the left-most column represent single variables or combinations of variables that make a significant contribution to the final model M (bottom row). Inclusion of these symbols in the 'model terms' column indicates that the corresponding combination of variables is included in that model. The likelihood ratio (LR) χ^2 measures the difference in deviance relative to that of the null model (5153.8), using the formulae defined in the Methods ($n=3,727$ for $E=0$ plus $E=5$ expression scores combined). The Δ AIC, which is equal to $(LR \chi^2 - 2 \cdot d.f.)$, is a standard measure of whether an improvement in model quality exceeds that expected at random from a model with a given number of degrees of freedom (d.f.)⁵³⁻⁵⁵. Because many compositional, free energy, and other terms were considered for inclusion in the model, a factor of 100 was used to correct for multiple-hypothesis testing, and parameters were only included in the final model M if significant at a Bonferroni-corrected false discovery rate of 5% (that is, $P \leq 0.05/100 = 5 \times 10^{-4}$). Most parameters are defined in the main text. The variables r and r_{codon} represent the amino acid and codon repetition rates, respectively, which were calculated as described in the Methods. The variables a_p , g_p and u_p indicate the fractional content of the respective base at the p position in codons throughout the gene. The subscripts 1-6, 1-16, 7-end, 7-16, 17-32, N-half and C-half indicate models in which the coefficients for the indicated parameter were optimized separately in the indicated range of codons or in the first or second halves of each gene, respectively. $\langle G_T \rangle_{96}$ is the average value of the predicted energy of mRNA folding in 50% overlapping windows 96 nucleotides long in the tail of the gene; although this parameter is strongly correlated with outcome (Fig. 1h), it does not appear in the final model, indicating that its apparent influence probably derives from mechanistic effects exerted by cross-correlated parameters (for example, as shown in Extended Data Fig. 3). 'Codons' indicates independent coefficients for the frequencies of each of the relevant 3-mers or 4-mers with one omitted to ensure proper normalization. The 'codons in frame' model excludes the three stop codons and AUG, effectively causing the contribution of this last codon to be absorbed into the constant term in the mathematical model. All calculations excluded the C-terminal LEHHHHHH tag carried by every protein. **b**, Effects of adding terms to the final computational model M. The LR χ^2 column shows the change in this value if the indicated term is added to model M, which is defined in the final row in **a**. Parameters and subscripts appearing in that table have the same definition here. SD_{ave} and SD_{max} represent the average and the maximum of a Boltzmann-weighting term for hybridization of sub-sequences in any frame to the anti-Shine-Dalgarno sequence at the 5' terminus of ribosomal 16S RNA, while f_{AGGA} represents the frequency in any frame of this subsequence that is complementary to its core. CAI, codon adaptation index¹⁴; tAI, tRNA adaptation index¹⁶.

Extended Data Table 2 | Codons used for synonymous gene design

Amino acid	<i>Number or identity of allowed codons</i>		
	Native genes	6AA Method	31C Method
Ala	4	4	GCT, GCA
Arg	6	CGT	CGT, CGA
Asn	2	2	AAT
Asp	2	GAT	GAT
Cys	2	2	TGT
Gln	2	CAA	CAA, CAG
Glu	2	GAA	GAA
Gly	4	4	GGT
His	2	CAT	CAT, CAC
Ile	3	ATT	ATT, ATC
Leu	6	6	TTA, TTG, CTA
Lys	3	3	AAA
Met	1	1	ATG
Phe	2	2	TTT
Pro	4	4	CCT, CCA
Ser	6	6	AGT, TCA
Thr	4	4	ACA, ACT
Trp	1	1	TGG
Tyr	2	2	TAT
Val	4	4	GTT, GTA

In our design of synonymous sequences, we reduced the native degeneracy of the genetic code to eliminate codons that correlate with reduced protein expression in single-variable logistic-regression analyses of our large-scale data set (dark grey symbols in Fig. 3a). In the 6AA approach, a single codon was used for 6 amino acids, while codons for the other 14 amino acids were not changed from the wild-type gene sequence. In the 31C-FO (and 31C-FD) approaches, the free energy was optimized (or de-optimized) using only the indicated subset of codons for each amino acid.

A prevalence of dynamo-generated magnetic fields in the cores of intermediate-mass stars

Dennis Stello^{1,2}, Matteo Cantiello³, Jim Fuller^{3,4}, Daniel Huber^{1,2,5}, Rafael A. García⁶, Timothy R. Bedding^{1,2}, Lars Bildsten^{3,7} & Victor Silva Aguirre²

Magnetic fields play a part in almost all stages of stellar evolution¹. Most low-mass stars, including the Sun, show surface fields that are generated by dynamo processes in their convective envelopes^{2,3}. Intermediate-mass stars do not have deep convective envelopes⁴, although 10 per cent exhibit strong surface fields that are presumed to be residuals from the star formation process⁵. These stars do have convective cores that might produce internal magnetic fields⁶, and these fields might survive into later stages of stellar evolution, but information has been limited by our inability to measure the fields below the stellar surface⁷. Here we report the strength of dipolar oscillation modes for a sample of 3,600 red giant stars. About 20 per cent of our sample show mode suppression, by strong magnetic fields in the cores⁸, but this fraction is a strong function of mass. Strong core fields occur only in red giants heavier than 1.1 solar masses, and the occurrence rate is at least 50 per cent for intermediate-mass stars (1.6–2.0 solar masses), indicating that powerful dynamos were very common in the previously convective cores of these stars.

Red giants are formed when a low- or intermediate-mass star has finished burning the hydrogen in its core. This leaves an inert helium core surrounded by a thin hydrogen-burning shell and a very thick outer convective envelope. Like the Sun, red giants oscillate in a broad comb-like frequency spectrum of radial and non-radial acoustic modes that

are excited by the turbulent surface convection⁹. The observed power spectrum has a roughly Gaussian envelope whose central frequency, ν_{\max} , decreases as a star expands during the red giant phase¹⁰. The comb structure of the spectrum arises from a series of overtone modes separated by the so-called large frequency separation, $\Delta\nu$. One of these overtone sequences is seen for each spherical degree, ℓ . For observations of unresolved distant stars, geometric cancellation prevents the detection of modes with $\ell > 3$. Their spectra are characterized by a pattern of radial ($\ell = 0$) and quadrupolar ($\ell = 2$) modes that form close pairs, interspersed with dipolar ($\ell = 1$) modes located roughly halfway between successive radial-quadrupolar pairs. The octupolar modes ($\ell = 3$) are weak or undetectable. The dipolar modes have turned out to be particularly useful probes of internal structure¹¹. They have been used to distinguish between hydrogen-shell and helium-core burning stars^{12–14} and to measure radial differential rotation^{15,16}. This is because each acoustic non-radial mode in the envelope couples to multiple gravity modes in the core, forming several observable mixed modes with frequencies in the vicinity of the acoustic mode¹⁵. This coupling is strongest for dipole modes, making them the most useful probes of the core¹⁷.

Figure 1 shows the oscillation power spectra of red giants at three different evolutionary stages observed by NASA's Kepler mission. For 'normal' stars (upper panels in Fig. 1), the dipolar modes (red peaks) have

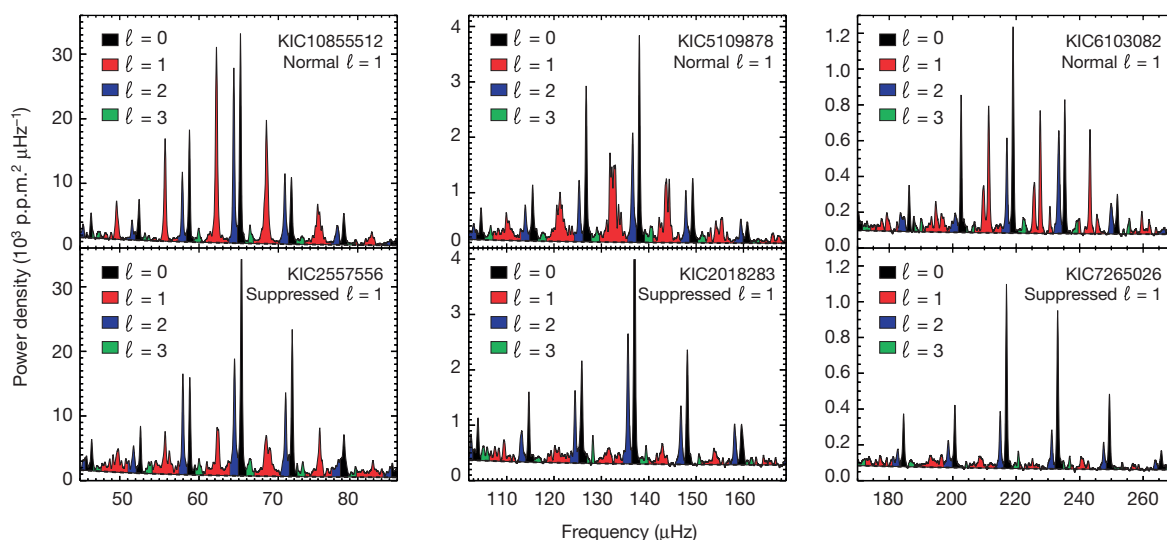


Figure 1 | Oscillation spectra of six red giants observed with Kepler. The stars are grouped into three pairs, each representing a different evolution stage ranging from the most evolved (lowest ν_{\max}) on the left to the least evolved (highest ν_{\max}) to the right. The coloured regions mark the power density dominated by modes of different degree $\ell = 0$ –3. For clarity the

spectra are smoothed by $0.03\Delta\nu$, which for the most evolved stars tends to create a single peak at each acoustic resonance, although each peak comprises multiple closely spaced mixed modes (red peaks in the left and centre panels). The slightly downward-sloping horizontal dashed line indicates the noise level.

¹Sydney Institute for Astronomy (SfA), School of Physics, University of Sydney, Sydney, New South Wales 2006, Australia. ²Stellar Astrophysics Centre, Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark. ³Kavli Institute for Theoretical Physics, University of California, Santa Barbara, California 93106, USA. ⁴TAPIR, Walter Burke Institute for Theoretical Physics, Mailcode 350-17, California Institute of Technology, Pasadena, California 91125, USA. ⁵SETI Institute, 189 Bernardo Avenue, Mountain View, California 94043, USA. ⁶Laboratoire AIM, CEA/DSM—CNRS—Université Paris Diderot—IRFU/Sap Centre de Saclay, 91191 Gif-sur-Yvette Cedex, France. ⁷Department of Physics, University of California, Santa Barbara, California 93106, USA.

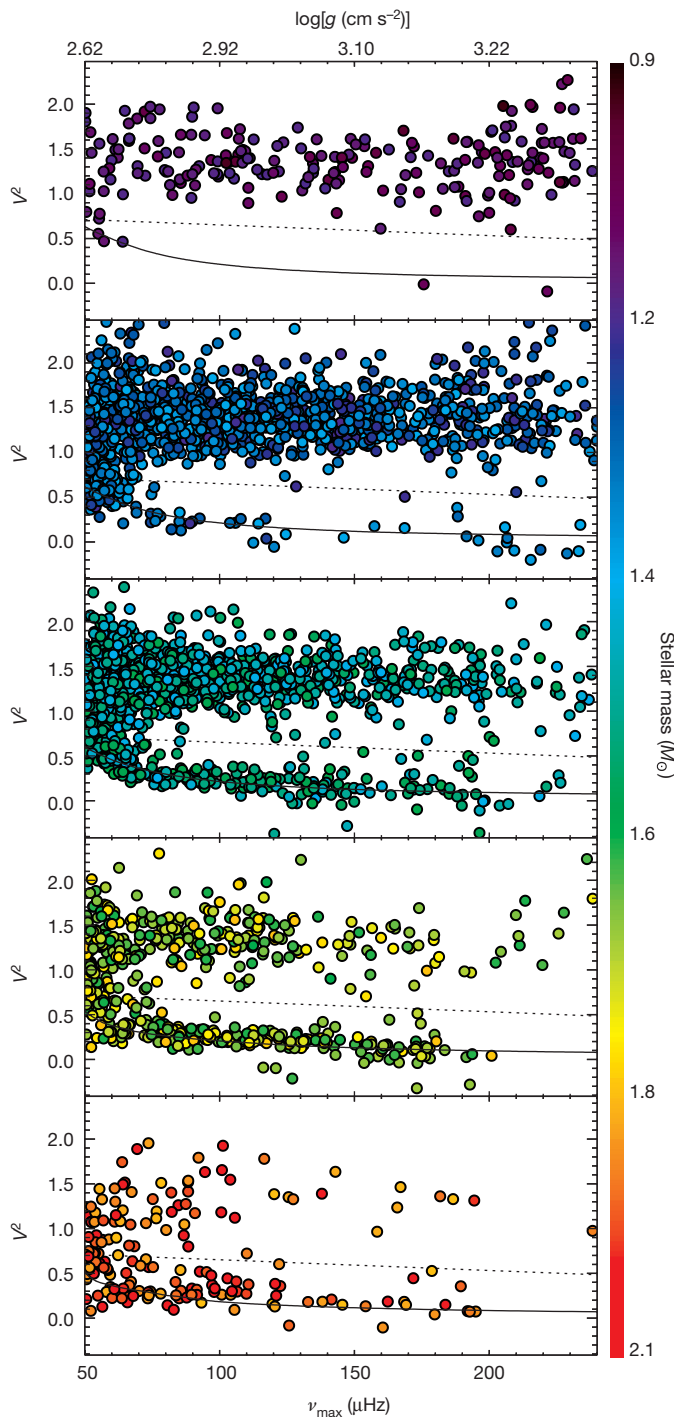


Figure 2 | Visibility of dipolar modes for red giants observed with Kepler. The abscissa shows the central frequency of each oscillation, which correlates closely with surface gravity (shown on the top axis). Stars evolve from right to left in the diagram, corresponding roughly to the beginning of the red giant phase to the red giant luminosity bump²⁷. Stellar mass¹³ (see colour scale on the right) has a formal 1σ uncertainty of 10% (ref. 28). The solid black line shows the theoretical predicted dipole-mode suppression⁸ for $1.1M_{\odot}$, $1.3M_{\odot}$, $1.5M_{\odot}$, $1.7M_{\odot}$ and $1.9M_{\odot}$ (top to bottom panels) and a radial-mode lifetime of 20 days²⁹. The fiducial dashed line separates normal and dipole-suppressed stars.

similar power to the radial modes (black peaks). However, at each stage of evolution we also find stars with greatly suppressed dipolar modes (lower panels in Fig. 1). Suppressed dipolar modes have been reported in a few dozen red giant stars^{18,19}, with an occurrence rate of about 20%.

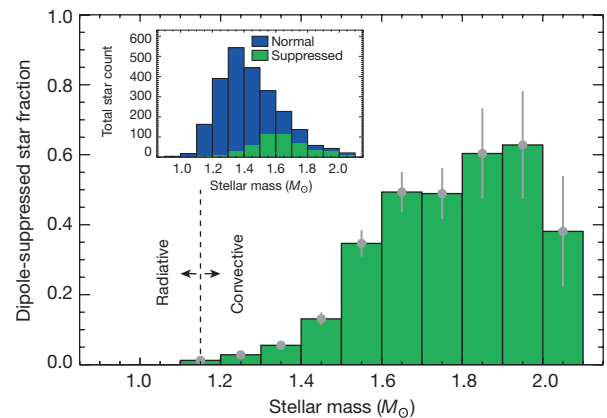


Figure 3 | Observed fraction of stars with suppressed dipolar modes.

The abscissa is the stellar mass (in multiples of solar mass, M_{\odot}). Dipole-suppressed stars are defined as those that fall below the dashed line in Fig. 2. Only stars with $\nu_{\max} > 70 \mu\text{Hz}$ are counted, to make the distinction between normal and suppressed unambiguous. The 1σ uncertainty in the fractions (grey vertical error bars) are based on Poisson statistics of the star counts (blue and green bars in the inset). The vertical dotted line separates stars for which hydrogen-core burning took place in either a radiative or convective environment for solar metallicity⁴.

The cause of this phenomenon has been puzzling until recent theoretical work⁸, which showed that the suppression can be explained if waves entering the stellar core are prevented from returning to the envelope. This occurs for dipolar modes if there are strong magnetic fields in the core, giving rise to a “magnetic greenhouse effect”⁸.

We measured the amount of suppression by comparing the integrated power of the dipolar and radial modes (the dipole mode visibility, V^2), averaged over the four orders centred on ν_{\max} (see Methods for details). While the normal stars show dipole mode visibilities of $V^2 \approx 1.5$, independently of ν_{\max} , the stars with suppressed modes have $V^2 \approx 0.5$ for $\nu_{\max} \approx 70 \mu\text{Hz}$ and down to almost zero for the least-evolved red giants oscillating above $200 \mu\text{Hz}$ (Fig. 1).

In Fig. 2 we show the dipole mode visibility for about 3,600 red giants observed over the first 37 months of the Kepler mission (see Methods). Our analysis is restricted to a sample of stars with ν_{\max} larger than $50 \mu\text{Hz}$ and masses below $2.1M_{\odot}$ (where M_{\odot} is the mass of the Sun), which, assuming no observational uncertainties, is expected to include only red giants that have not started burning helium in their cores¹³. We cross-matched our sample with those of known helium-burning stars^{13,14}, which allowed us to identify and remove a small fraction of evolved stars burning helium that, owing to measurement uncertainty, had entered our sample (2% of our sample, almost all with $\nu_{\max} < 70 \mu\text{Hz}$).

The stars in Fig. 2 form two distinct branches that gradually merge as the stars evolve leftwards towards lower ν_{\max} . Most stars fall on the ‘normal’ upper branch of $V^2 \approx 1.5$, in agreement with previous results¹⁸. The lower branch, with suppressed dipole modes, agrees remarkably well with theoretical predictions (black curve). This prediction assumes that all the wave energy leaking into the stellar core is trapped by a magnetic greenhouse effect caused by strong internal magnetic fields⁸. The decrease of the suppression towards lower ν_{\max} is a consequence of the weaker coupling between acoustic waves in the envelope and gravity waves in the core⁸. With this large sample we have been able to separate the stars in Fig. 2 into five different mass intervals, from $0.9M_{\odot}$ to $2.1M_{\odot}$. It is striking how strongly the relative population on the lower branch (stars with suppressed dipole modes) depends on mass.

We quantify the mass dependence in Fig. 3 by showing the relative number of dipole-suppressed stars (those below the dashed line in Fig. 2) in narrow mass intervals. We see no suppression in red giants below $1.1M_{\odot}$, which coincides with the mass below which they did not have convective cores during the core-hydrogen-burning phase⁴.

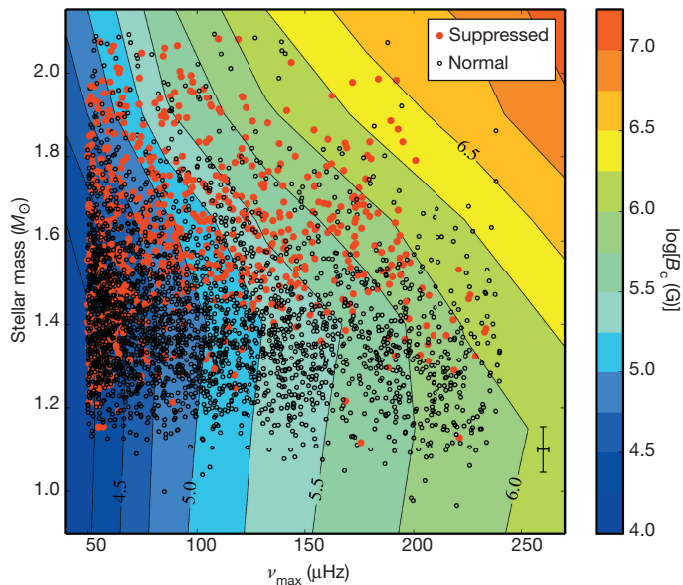


Figure 4 | Critical magnetic field strength required to suppress dipole mode oscillations. The abscissa is the observed central frequency of the oscillations. The ordinate is the inferred asteroseismic mass. The coloured contours indicate the minimum magnetic field at the hydrogen shell required for mode suppression (the critical field, B_c). Filled red circles mark stars with observed suppressed modes, and open circles mark normal (non-suppressed) stars. The cross shows a typical 1σ error bar for the data. The uncertainty in B_c due to uncertainty in mass is negligible for stars below $1.4M_\odot$ and is below 25% for the more massive stars.

The onset of magnetic suppression above this threshold suggests that at least some of those stars had convectively driven magnetic dynamos in their cores during the core-hydrogen-burning (main-sequence) phase. This is supported by three-dimensional hydrodynamical modelling of these stars⁶. Red giants no longer contain convective cores, leading us to conclude that the strong magnetic fields in suppressed oscillators are the remnants of the fields produced by core dynamos during the main sequence.

Figure 3 shows that the incidence of magnetic suppression increases with mass, with red giants above $1.6M_\odot$ showing a remarkable suppression rate of 50% to 60%. These have evolved from main-sequence A-type stars, among which only up to about 10% are observed to have strong fields at their surfaces⁵. We conclude that these magnetic A-type stars represent only the tip of the iceberg, and that a much larger fraction of A-type stars have strong magnetic fields hidden in their cores.

In Fig. 4 we show the observed ν_{\max} and inferred mass of all the stars superimposed on a contour plot of minimum magnetic field strengths required for mode suppression⁸. For stars with suppressed modes (filled red circles), the underlying colour provides a lower bound to the field strength at the hydrogen-burning shell. For stars without suppressed modes (open black circles), the underlying colour represents an upper limit to the field at the hydrogen-burning shell; above or below the shell the field could potentially be larger. Hence, normal and dipole-suppressed stars that fall in the same regions of Fig. 4 may have core field strengths that are only slightly different. However, we expect that the dipole-suppressed stars on average exhibit stronger core fields than their normal counter parts.

Considering again the low-mass stars ($<1.1M_\odot$), of which none show suppression, we see from Fig. 4 that radial magnetic fields above about 100 kG are not present at the hydrogen-burning shell when the stars are just below the red giant luminosity bump ($\nu_{\max} \approx 70\text{--}100\mu\text{Hz}$). Assuming magnetic flux conservation from the main-sequence phase, this suggests that radial fields above approximately 5 kG do not exist within the cores of Sun-like stars⁸. Large-scale fields in the solar interiors have been suggested to explain the properties of the tachocline²⁰. However, our results do not rule out strong horizontal fields near the

radiative–convective boundary because those fields would be outside the core and could not cause mode suppression when the star evolves into a red giant.

Turning to higher masses we see that, for a given ν_{\max} , stars above $1.4M_\odot$ require increasingly strong magnetic fields to suppress their dipolar modes. From Fig. 4, there is no clear upper limit to the field strengths present in red giant cores, given that dipole-suppressed stars are common even when field strengths $B > 1$ MG are required for suppression. However, the hint of a decline in the occurrence of dipole-suppressed stars above $2M_\odot$ seen in Fig. 3 suggests there may be a mass above which dynamo-generated magnetic fields are less likely to cause oscillation mode suppression in intermediate-mass stars.

The high occurrence rate of dipole mode suppression demonstrates that internal magnetic fields, generated by a convective core dynamo during the main sequence, can persist through the red giant phase. This indicates that these dynamo-generated fields are frequently able to settle into long-lived stable configurations^{21,22}. The occurrence rate of suppressed dipole modes in intermediate-mass red giants is much higher than the occurrence rate of strong fields at the surfaces of the main-sequence A-type stars from which they evolved. These surface fields are thought to be a relic of the star's formation process^{21,23}. We conclude that fields generated during convective hydrogen-core burning are able to settle into stable equilibrium configurations much more often (more than 50% of the time) than fields generated or inherited during star formation (less than 10% of the time).

Our results show that main-sequence stars with no observable magnetic field at the surface can still harbour strong fields in the core that survive into the red giant phase. The presence of internal magnetic fields might play an important part in angular momentum transport and potentially influence internal mixing processes. Fields too weak to suppress dipolar oscillation modes may exist in normal red giants, but these fields may nevertheless transport enough angular momentum to help explain the measured rotation rates of red giant cores^{16,24}. The presence of strong internal magnetic fields might help to explain the extra internal mixing observed in red giant branch stars and asymptotic giant branch stars²⁵. Interestingly, the magnetic field strength necessary for magnetic buoyancy mixing²⁶ is similar to the field strength required for dipole mode suppression⁸.

After some time, intermediate-mass red giants also start burning helium in their cores. Suppressed dipolar modes in those so-called red clump stars will reveal whether the fields survive until helium-core burning, and whether they can account for magnetic fields observed in stellar remnants such as white dwarfs. Like intermediate-mass stars, more massive stars ($>10M_\odot$) also undergo convective hydrogen-core burning that generates a magnetic dynamo, and which may produce the magnetic fields observed in many neutron stars.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 June; accepted 15 October 2015.

Published online 4 January 2016.

- Landstreet, J. D. Magnetic fields at the surfaces of stars. *Astron. Astrophys. Rev.* **4**, 35–77 (1992).
- Parker, E. N. Hydromagnetic dynamo models. *Astrophys. J.* **122**, 293–314 (1955).
- Donati, J.-F. & Landstreet, J. Magnetic fields of nondegenerate stars. *Annu. Rev. Astron. Astrophys.* **47**, 333–370 (2009).
- Kippenhahn, R. & Weigert, A. *Stellar Structure and Evolution* (Springer, 1990).
- Power, J., Wade, G. A., Aurière, M., Silvester, J. & Hanes, D. Properties of a volume-limited sample of Ap-stars. *Contrib. Astron. Observ. Skalnaté Pleso* **38**, 443–444 (2008).
- Brun, A. S., Browning, M. K. & Toomre, J. Simulations of core convection in rotating A-type stars: magnetic dynamo action. *Astrophys. J.* **629**, 461–481 (2005).
- Aurière, M. *et al.* The magnetic fields at the surface of active single G-K giants. *Astron. Astrophys.* **574**, A90 (2015).
- Fuller, J., Cantiello, M., Stello, D., Garcia, R. A. & Bildsten, L. Strong internal magnetic fields explain suppressed oscillation modes in red giant stars. *Science* **350**, 423–426 (2015).

9. De Ridder, J. D. *et al.* Non-radial oscillation modes with long lifetimes in giant stars. *Nature* **459**, 398–400 (2009).
10. Stello, D., Bruntt, H., Preston, H. & Buzasi, D. Oscillating K giants with the *WIRE* satellite: determination of their asteroseismic masses. *Astrophys. J.* **674**, L53–L56 (2008).
11. García, R. A. & Stello, D. in *Extraterrestrial Seismology* (eds Tong, V. C. H. & García, R. A.) Ch. 11 (Cambridge Univ. Press, 2015).
12. Bedding, T. R. *et al.* Gravity modes as a way to distinguish between hydrogen- and helium-burning red giant stars. *Nature* **471**, 608–611 (2011).
13. Stello, D. *et al.* Asteroseismic classification of stellar populations among 13,000 red giants observed by Kepler. *Astrophys. J.* **765**, L41 (2013).
14. Mosser, B. *et al.* Mixed modes in red giants: a window on stellar evolution. *Astron. Astrophys.* **572**, L5 (2014).
15. Beck, P. G. *et al.* Kepler detected gravity-mode period spacings in a red giant star. *Science* **332**, 205 (2011).
16. Mosser, B. *et al.* Spin down of the core rotation in red giants. *Astron. Astrophys.* **548**, A10 (2012).
17. Dupret, M.-A. *et al.* Theoretical amplitudes and lifetimes of non-radial solar-like oscillations in red giants. *Astron. Astrophys.* **506**, 57–67 (2009).
18. Mosser, B. *et al.* Characterization of the power excess of solar-like oscillations in red giants with Kepler. *Astron. Astrophys.* **537**, A30 (2012).
19. García, R. A. *et al.* Study of KIC 8561221 observed by Kepler: an early red giant showing depressed dipolar modes. *Astron. Astrophys.* **563**, A84 (2014).
20. Gough, D. O. & McIntyre, M. E. Inevitability of a magnetic field in the Sun's radiative interior. *Nature* **394**, 755–757 (1998).
21. Braithwaite, J. & Spruit, H. C. A fossil origin for the magnetic field in A stars and white dwarfs. *Nature* **431**, 819–821 (2004).
22. Duez, V., Braithwaite, J. & Mathis, S. On the stability of non-force-free magnetic equilibria in stars. *Astrophys. J.* **724**, L34–L38 (2010).
23. Moss, D. On the magnetic flux distribution in magnetic CP stars. *Mon. Not. R. Astron. Soc.* **226**, 297–307 (1987).
24. Cantiello, M., Mankovich, C., Bildsten, L., Christensen-Dalsgaard, J. & Paxton, B. Angular momentum transport within evolved low-mass stars. *Astrophys. J.* **788**, 93 (2014).
25. Busso, M., Wasserburg, G. J., Nollett, K. M. & Calandra, A. Can extra mixing in RGB and AGB stars be attributed to magnetic mechanisms? *Astrophys. J.* **671**, 802–810 (2007).
26. Nucci, M. C. & Busso, M. Magnetohydrodynamics and deep mixing in evolved stars. I. Two- and three-dimensional analytical models for the asymptotic giant branch. *Astrophys. J.* **787**, 141 (2014).
27. Salaris, M. & Cassisi, S. *Evolution of Stars and Stellar Populations* (John Wiley & Sons, 2005).
28. Miglio, A. *et al.* Asteroseismology of old open clusters with Kepler: direct estimate of the integrated red giant branch mass-loss in NGC 6791 and 6819. *Mon. Not. R. Astron. Soc.* **419**, 2077–2088 (2012).
29. Corsaro, E., Ridder, J. D. & García, R. A. Bayesian peak bagging analysis of 19 low-mass low-luminosity red giants observed with Kepler. *Astron. Astrophys.* **579**, A83 (2015).

Acknowledgements This paper has been written collaboratively, on the web, using Authorea (<http://www.authorea.com>). We acknowledge the entire Kepler team, whose efforts made these results possible. D.S. is the recipient of an Australian Research Council Future Fellowship (project number FT140100147). J.F. acknowledges support from NSF under grant number AST-1205732 and through a Lee DuBridge Fellowship at Caltech. R.A.G. acknowledges the support of the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement number 269194 (IRSES/ASK), the CNES, and the ANR-12-BS05-0008, IDEE. D.H. acknowledges support by the Australian Research Council's Discovery Projects funding scheme (project number DE140101364) and support by the National Aeronautics and Space Administration under grant number NNX14AB92G issued through the Kepler Participating Scientist Program. This project was supported by NASA under TCAN grant number NNX14AB53G, and by the NSF under grant numbers PHY11-25915 and AST11-09174. Funding for the Stellar Astrophysics Centre is provided by The Danish National Research Foundation (grant agreement number DNRF106). The research is supported by the ASTERISK project (ASTERoseismic Investigations with SONG and Kepler) funded by the European Research Council (grant agreement number 267864).

Author Contributions D.S. measured and interpreted mode visibilities; M.C. and J.F. calculated and interpreted theoretical models; D.H. and D.S. calculated power spectra and measured large frequency separations; R.A.G., T.R.B., L.B. and V.S.A. contributed to the discussion of the results. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.S. (stello@physics.usyd.edu.au).

METHODS

Dipole mode visibilities. For each star in our sample we derived the frequency power spectrum as the Fourier transform of the Kepler light curve (up to observing quarter 14) and adopted the values of $\Delta\nu$, ν_{\max} and mass from previous work¹³. To correct the spectrum for the background noise, we measured and subtracted a linear slope anchored on the median power on either side of the central power excess, defined by the frequency ranges $0.75(\nu_{\max} - 2\Delta\nu)$ to $(\nu_{\max} - 2\Delta\nu)$ and $(\nu_{\max} + 2\Delta\nu)$ to $1.25(\nu_{\max} + 2\Delta\nu)$. We then selected a $4\Delta\nu$ -wide range of the spectrum centred on ν_{\max} . The location of each mode was found by first folding this central part of the spectrum using $\Delta\nu$ as the folding frequency such that modes of the same spherical degree each formed a single peak. The folded spectrum was smoothed by a Gaussian filter with a width of $0.1\Delta\nu$. We finally correlated the folded spectrum with a model spectrum comprised of three Lorentzian profiles, one for each degree $\ell = 0, 1$ and 2 , with relative heights of $1.0, 0.5$ and 0.8 , and widths of $5\%, 10\%$ and 5% of $\Delta\nu$, respectively. The centres of each Lorentzian profile were fixed relative to one another such that the one representing the $\ell = 2$ modes was $0.12\Delta\nu$ to the left (lower frequency) of the $\ell = 0$ profile, and the $\ell = 1$ profile was $0.52\Delta\nu$ to the

right (higher frequency)³⁰. The shift between the model and the observed folded spectrum that gave the largest correlations provided the location of each mode. The regions of the power spectrum associated with each spherical degree were set to be $0.16\Delta\nu, 0.53\Delta\nu$ and $0.16\Delta\nu$ wide for $\ell = 0-2$, respectively, with the regions located according to the mode location found by the correlation with the model spectrum. The remaining region was associated with $\ell = 3$ modes (Fig. 1). The dipole-mode visibility was derived as the integrated power of the dipole modes relative to that of the radial modes, following the approach by previous studies¹⁸.

Code availability. We used the commercially available IDL routine *lmp_test* to derive the Fourier spectra of the Kepler light curves. To derive the mode visibilities we used built-in IDL routines and subroutines written by others, which are therefore not made publicly available. For the stellar evolution calculations we used MESA (publicly available at <http://mesa.sourceforge.net/>).

30. Huber, D. *et al.* Asteroseismology of red giants from the first four months of Kepler data: global oscillation parameters for 800 stars. *Astrophys. J.* **723**, 1607–1617 (2010).

Exposed water ice on the nucleus of comet 67P/Churyumov–Gerasimenko

G. Filacchione¹, M. C. De Sanctis¹, F. Capaccioni¹, A. Raponi¹, F. Tosi¹, M. Ciarniello¹, P. Cerroni¹, G. Piccioni¹, M. T. Capria¹, E. Palomba¹, G. Bellucci¹, S. Erard², D. Bockelee–Morvan², C. Leyrat², G. Arnold³, M. A. Barucci², M. Fulchignoni², B. Schmitt⁴, E. Quirico⁴, R. Jaumann³, K. Stephan³, A. Longobardo¹, V. Mennella⁵, A. Migliorini¹, E. Ammannito⁶, J. Benkhoff⁷, J. P. Bibring⁸, A. Blanco⁹, M. I. Blecka¹⁰, R. Carlson¹¹, U. Carsenty³, L. Colangeli⁷, M. Combes², M. Combi¹², J. Crovisier², P. Drossart², T. Encrenaz², C. Federico¹³, U. Fink¹⁴, S. Fonti⁹, W. H. Ip¹⁵, P. Irwin¹⁶, E. Kuehrt³, Y. Langevin⁸, G. Magni¹, T. McCord¹⁷, L. Moroz³, S. Mottola³, V. Orofino⁹, U. Schade¹⁸, F. Taylor¹⁶, D. Tiphene², G. P. Tozzi¹⁹, P. Beck⁴, N. Biver², L. Bonal⁴, J–Ph. Combe¹⁷, D. Despan², E. Flamini²⁰, M. Formisano¹, S. Fornasier², A. Frigeri¹, D. Grassi¹, M. S. Gudipati¹¹, D. Kappel³, F. Mancarella⁹, K. Markus³, F. Merlin², R. Orosei²¹, G. Rinaldi¹, M. Cartacci¹, A. Cicchetti¹, S. Giuppi¹, Y. Hello², F. Henry², S. Jacquino², J. M. Reess², R. Noschese¹, R. Politi¹ & G. Peter²²

Although water vapour is the main species observed in the coma of comet 67P/Churyumov–Gerasimenko^{1,2} and water is the major constituent of cometary nuclei^{3,4}, limited evidence for exposed water-ice regions on the surface of the nucleus has been found so far^{5,6}. The absence of large regions of exposed water ice seems a common finding on the surfaces of many of the comets observed so far^{7–9}. The nucleus of 67P/Churyumov–Gerasimenko appears to be fairly uniformly coated with dark, dehydrated, refractory and organic-rich material¹⁰. Here we report the identification at infrared wavelengths of water ice on two debris falls in the Imhotep region of the nucleus. The ice has been exposed on the walls of elevated structures and at the base of the walls. A quantitative derivation of the abundance of ice in these regions indicates the presence of millimetre-sized pure water-ice grains, considerably larger than in all previous observations^{6,7–9}. Although micrometre-sized water-ice grains are the usual result of vapour recondensation in ice-free layers⁶, the occurrence of millimetre-sized grains of pure ice as observed in the Imhotep debris falls is best explained by grain growth by vapour diffusion in ice-rich layers, or by sintering. As a consequence of these processes, the nucleus can develop an extended and complex coating in which the outer dehydrated crust¹⁰ is superimposed on layers enriched in water ice. The stratigraphy observed on 67P/Churyumov–Gerasimenko^{11,12} is therefore the result of evolutionary processes affecting the uppermost metres of the nucleus and does not necessarily require a global layering to have occurred at the time of the comet's formation.

Located in the bottom part of the main lobe of 67P/Churyumov–Gerasimenko, the Imhotep plain is characterized by a smooth terrain unit 1 km × 1.5 km wide filled by very fine dust^{11,13}. The region is circumscribed by elevated structures, some of them showing a circular shape with well defined walls, which display erosion and mass wasting on their sides¹⁴. Rosetta's navigation cameras identified bright albedo patches (BAPs) on the walls of two of these features (Fig. 1): a flat-floored raised circular structure with an external rim partially collapsed located at the Ash–Khepry–Imhotep boundary at longitude 117° E, latitude 13° N (BAP1) and a hill with a steep slope at longitude 180°–182° E, latitude 4°–10° S (BAP2). Both areas are exposed towards the lower Imhotep plain where the waste material is accumulated as boulders and debris

at the bottom of the cliffs. These deposits are located on relatively cold, collapsed walls of elevated structures and seem to be recently exposed. The debris falls indicate material eroded and deposited at the base of walls. The overall geological context points to a process in which limited amounts of bedrock material break out of the wall and are distributed as disrupted fragments at the bottom of the affected wall regions. Mechanisms that can trigger the falls are related to water-ice sublimation, water-ice metamorphism, thermal cracking or possibly impacts.

The best views of the two BAPs obtained by the Visual InfraRed Thermal Imaging Spectrometer (VIRTIS-M)¹⁵ are shown in Fig. 2. The description of these observations, including the retrieval of the spectral parameters and the associated temperature maps, are discussed in the Methods.

In the two BAP regions, VIRTIS-M has detected prominent absorption features at 1.05 μm, 1.25 μm, 1.5 μm (partial), and 2.0 μm (discussed later in Fig. 3). These spectral features represent clear evidence of the presence of water ice. When observing the water-ice-rich areas, the 3.2-μm band, ubiquitously observed throughout the cometary nucleus and interpreted¹⁰ to be caused by the presence of organic compounds, appears to be deformed by the strong 3.0-μm signature of water ice. This is the case for the Hapi active region, where it appears in the form of transient deposits formed in the upper layer of the surface by recondensation of water molecules sublimated from sub-surface ice during the diurnal cycle⁶.

With the aim of verifying the spectral properties of the observed water-ice regions and to study their persistence during the first few months of the mission, we have performed a temporal analysis of the appearance of the 2-μm band while we use the entire 1–4-μm spectral range to derive the water-ice abundance and grain size quantitatively. In Fig. 3 we show the 2-μm band depth time series (see Methods for analysis details) for a BAP1 neighbourhood using all the spectra measured in the period 23 September to 21 November 2014 (Fig. 3a, b) and for BAP2 obtained between 2 and 23 September 2014 (Fig. 3c). The persistence of the 2-μm band absorption is evident during these time periods, resulting in average values of 5.6%–7.6% for BAP1 and 5.0% for BAP2.

The spectral analyses show not much change in the band centres and depths during the period of observation, indicating that the ice is

¹INAF-IAPS, Istituto di Astrofisica e Planetologia Spaziali, Rome, Italy. ²LESIA, Observatoire de Paris/CNRS/UPMC/Université Paris-Diderot, Meudon, France. ³Institute for Planetary Research, DLR, Berlin, Germany. ⁴Université Grenoble Alpes, CNRS, IPAG, Grenoble, France. ⁵INAF-Osservatorio di Capodimonte, Napoli, Italy. ⁶UCLA, Los Angeles, California, USA. ⁷European Space Agency—ESTEC, Noordwijk, The Netherlands. ⁸Institut d'Astrophysique Spatial CNRS, Orsay, France. ⁹Dipartimento di Matematica e Fisica “Ennio De Giorgi”, Università del Salento, Lecce, Italy. ¹⁰Space Research Centre, Polish Academy of Sciences, Warsaw, Poland. ¹¹NASA JPL, Pasadena, California, USA. ¹²Space Physics Research Laboratory, The University of Michigan, Michigan, Ann Arbor, USA. ¹³Università di Perugia, Perugia, Italy. ¹⁴Lunar Planetary Laboratory, University of Arizona, Tucson, Arizona, USA. ¹⁵National Central University, Taipei, Taiwan. ¹⁶Department of Physics, Oxford University, Oxford, UK. ¹⁷Bear Fight Institute, Winthrop, Washington, USA. ¹⁸Helmholtz-Zentrum Berlin für Materialien und Energie, Berlin, Germany. ¹⁹INAF-Osservatorio Astrofisico di Arcetri, Firenze, Italy. ²⁰Agenzia Spaziale Italiana, Rome, Italy. ²¹Istituto di Radioastronomia—INAF, Bologna, Italy. ²²Institute of Optical Sensor Systems, DLR, Berlin, Germany.

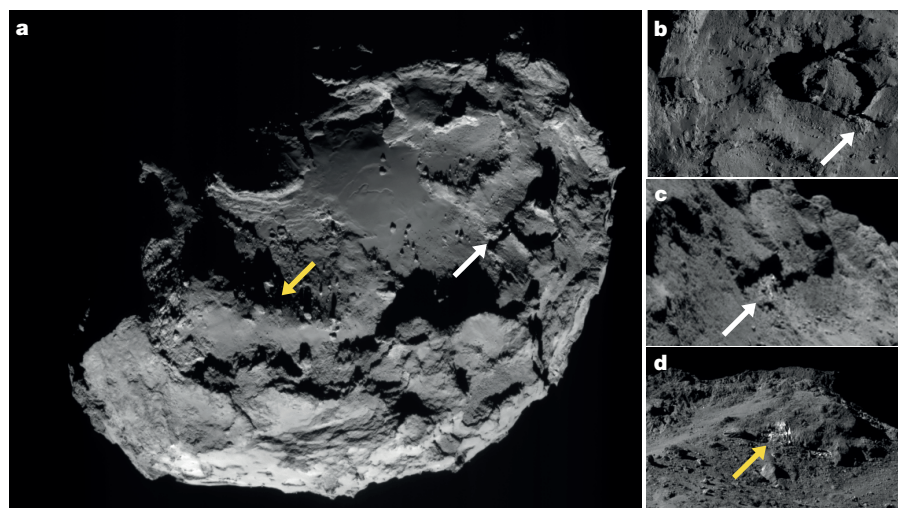


Figure 1 | Rosetta NAVCAM context images of the two debris falls. **a**, Imhotep region showing BAP1 on the Ash–Khepry–Imhotep boundary (white arrow) and BAP2 on the hillside (yellow arrow). **b**, Waste material and debris accumulated at the bottom of the landslide eroding the side of the elevated circular structure. BAP1 is not visible in the viewing geometry of this nadir-pointing image. **c**, BAP1 is caused by the presence of water ice, clearly visible on this oblique view of the elevated structure wall. **d**, Similar viewing geometry, identifying BAP2.

stable over that time frame. After linear spectral continuum removal, the 2.0- μm band centre appears to be located at 2.04 μm for BAP1 (Fig. 3d), with an average band depth of 5%. A band centre located at wavelengths slightly longer than 2.0 μm is diagnostic of crystalline water ice, as opposed to the amorphous ice phase, which has the band centre located at 2.0 μm (ref. 16). A secondary minimum of the 1.5- μm band, located at about 1.65 μm , is a typical feature of water ice in the crystalline form (it is absent in amorphous ice¹⁷); in Fig. 3e, the VIRTIS-M data partially show the long-wavelength wing of the 1.5- μm band, in which the secondary minimum at about 1.65 μm is clearly seen. Finally, in a few well resolved pixels the weaker 1.05- μm and 1.25- μm bands of water ice have also been identified (Fig. 3e).

VIRTIS-M reflectance spectra were modelled following the approach described in the Methods. The best-fitting result shown in Fig. 4a, for a representative spectrum of BAP1, corresponds to an areal

mixing in which 1.2% of the pixel area is occupied by patches of large grains of diameter around 2 mm for this specific case of pure water ice while the remaining 98.8% of the pixel area is composed of an intimate mixture of water ice (3.4%, grains of diameter 56 μm) and dark terrain (95.4%). As a result, the total amount of surface water ice on the simulated spectrum corresponds to about 4.6%, assuming each pixel samples an area of about 2.5 m \times 2.5 m. By repeating this fitting technique on each pixel of the two high-resolution images shown in Fig. 2c–e, we have derived water-ice abundance maps for the two debris falls in which intimate and areal mixing of ice coexist within the dark terrain (Fig. 4b–e). The maximum water-ice abundances, equivalent to about 1.2% and 4% for the areal and intimate cases respectively, correspond to those of the bluish waste material accumulated at the base of the two debris falls. A similar water-ice abundance (6%) was observed by the Deep Impact mission on the 0.5-km²-wide water ice

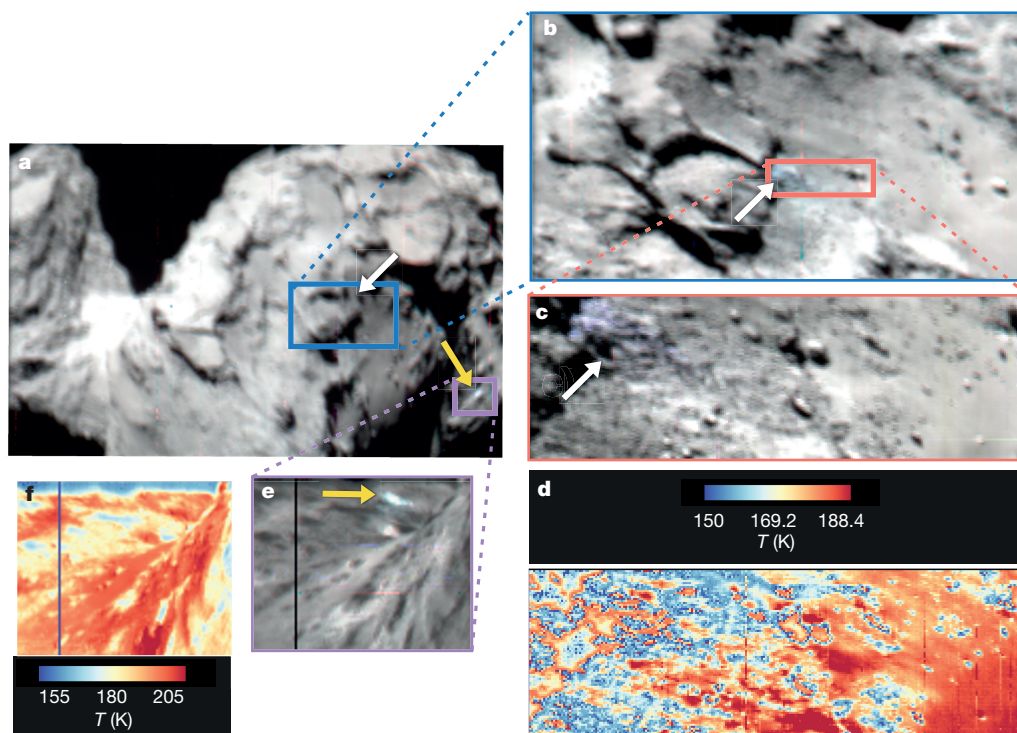


Figure 2 | VIRTIS-M observations of the two water-ice debris falls. **a**, Colour image showing BAP1 and BAP2 (white and yellow arrows, respectively). **b**, Water-ice-rich waste terrain on BAP1 is visible (blue) on the right side of the circular elevated structure. **c**, Close-up image acquisition of BAP1 reveals the bluish colour of the water-ice-rich unit.

d, Temperature map showing the water-ice-rich terrain at $T < 160$ K. **e**, Hillside image showing BAP2. **f**, Temperature map showing the water-ice-rich terrain at $T < 180$ K. All colour images are obtained from the $B = 1.3$ μm , $G = 2.0$ μm and $R = 2.9$ μm channels.

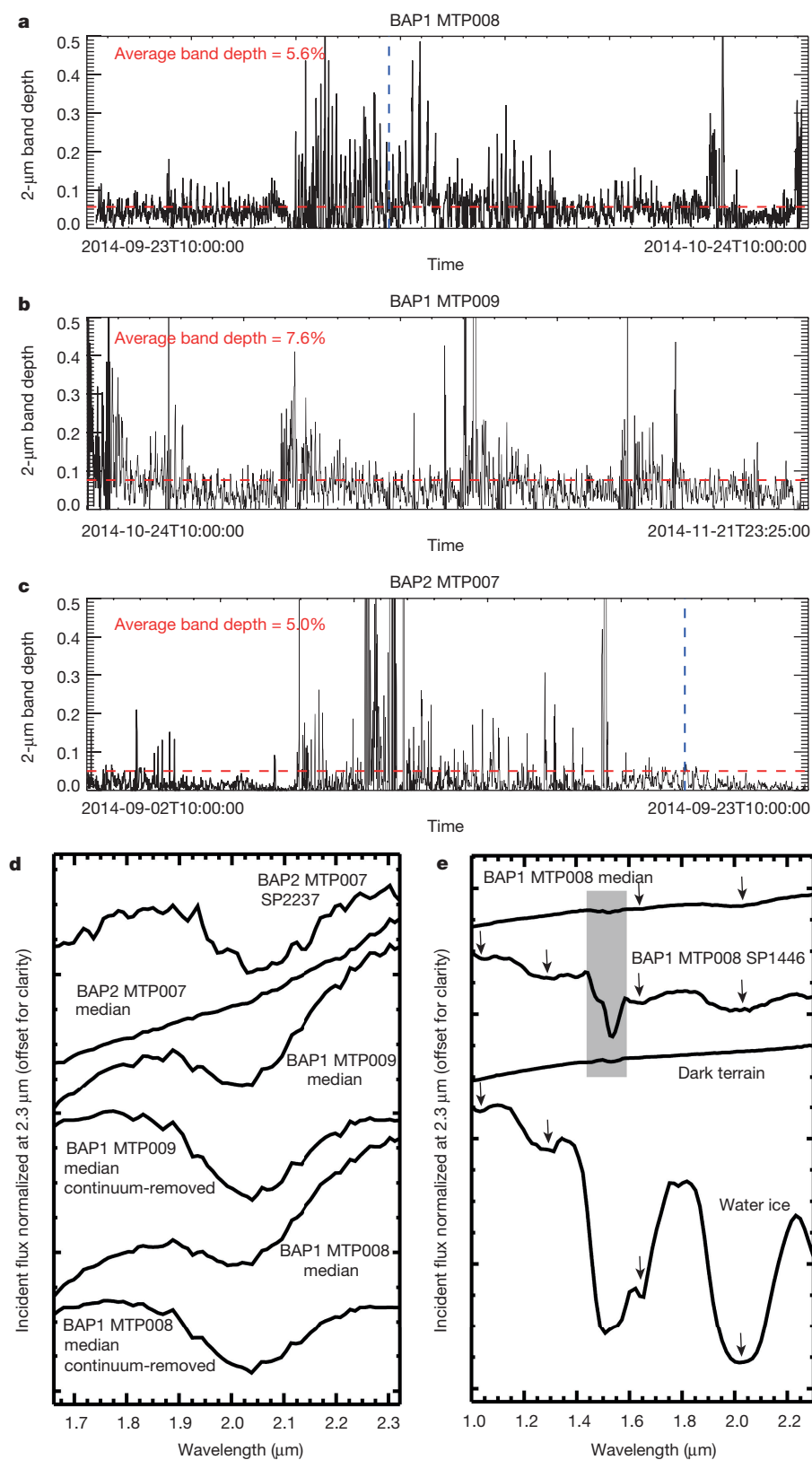


Figure 3 | Spectral evidence of water ice.

a, b, 2- μ m band depth time series for BAP1 in October (MTP008) and November (MTP009) 2014. **c**, Same as panels **a** and **b** but for BAP2 in September (MTP007) 2014. Fluctuations are caused by the varying signal-to-noise ratio, owing to the changing illumination and viewing conditions during the observations. **d**, Median spectra derived from the MTP008 and MTP009 data sets for BAP1, showing the 2- μ m absorption feature. Continuum-removed data show the spectral position of the band centre at about 2.04 μ m, compatible with water ice in crystalline form. The 2- μ m feature is not evident on the BAP2 median spectrum derived from the entire MTP007 data set; it appears only on limited observations, like the SP2237 one, marked by a blue vertical dashed line in **c**. SP2237 and SP1446 are the 2237th and 1446th spectra of the timeseries. **e**, The BAP1 median spectrum shows the 1.65- μ m and 2- μ m bands. A water-ice-rich pixel (SP1446, indicated by the blue dashed line in **a**) shows the 1.05- μ m, 1.25- μ m, 1.5–1.65- μ m and 2.05- μ m pure crystalline water-ice absorption features. This spectrum can be simulated with two endmembers, for example, crystalline water ice and dark terrain, as shown in Fig. 4. Instrumental order sorting filter junction wavelengths are marked by grey box.

deposits identified, by using a linear mixing technique, on the surface of comet 9P/Tempel 1 (ref. 7).

The size range of ice grains that produces the best fits in the observed areas displays a clear bimodal distribution (Extended Data Fig. 1), where 76% of the pixels in the intimate mixtures have grain sizes of 33–72 μ m, and 68% of the pixels in the pure ice regions have grain sizes of 1.4–2.6 μ m. Previously⁶, we showed that the diurnal mechanism of

ice trapping in the dehydrated uppermost layer led to the formation of ice grains only a few micrometres in diameter. Thus, the VIRTIS-M observations indicate the presence of three different populations of icy grains on the nucleus surface, which implies different growth mechanisms operating on different spatial scales and timescales.

Small water-ice grains are the result of the fast condensation process occurring for each comet rotation on the Hapi area in the thin

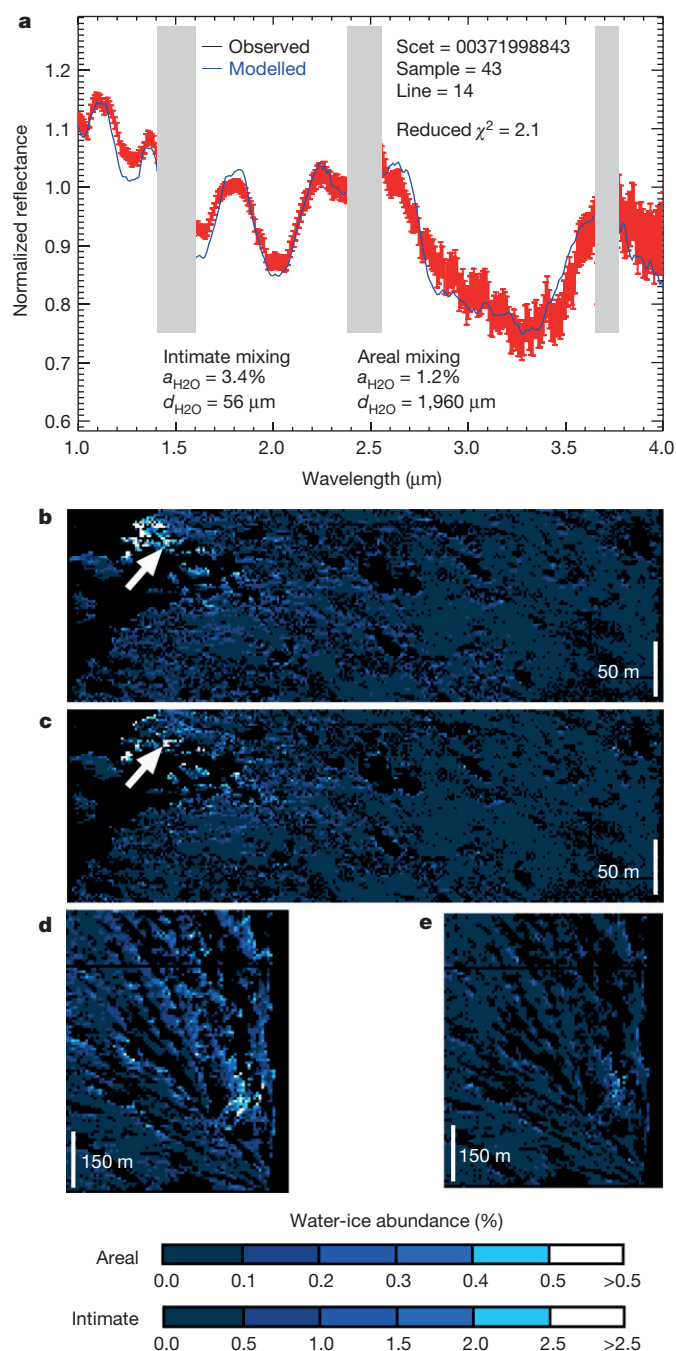


Figure 4 | Water-ice spectral analysis and spatial distribution.

a, BAP1 water-ice-rich VIRTIS spectrum (black curve with error bars in red; at each wavelength the error corresponds to the inverse of the signal-to-noise ratio) and best-fit synthetic model (blue curve). $a_{\text{H}_2\text{O}}$ is the percentage of water ice and $d_{\text{H}_2\text{O}}$ is the water-ice grain size. 'Scet' is the spacecraft time used to identify the observation; 'sample' and 'line' correspond to the position of the pixel on the image. The location of the VIRTIS spectrum corresponds to the pixel indicated by an arrow in **b** and **c**. Instrumental order sorting filter junction wavelengths are marked by grey boxes. **b**, **c**, Water-ice abundance maps for BAP1 on observation I1_00371998843 (Fig. 2c) in areal (**b**) and intimate (**c**) mixing with dark material, respectively. **d**, **e**, Water-ice abundance maps for BAP2 on observation I1_00369369214 (Fig. 2e) in areal (**d**) and intimate (**e**) mixing with dark material, respectively. Black pixels in **b–e** correspond mainly to shadow areas whose spectra are excluded from the analysis because they have a median signal-to-noise ratio <30 , and for which a satisfactory model fit was not obtained. The estimated relative error for retrieved parameters is 20%.

(centimetre-sized) uppermost ice-free layers⁶. This is consistent with Deep Impact observations of the size distribution of ice grains released immediately after the impact event¹⁸. On the other hand, the occurrence of millimetre-sized grains of 'pure' (that is, water) ice, as observed over large areas (metre-sized) on the Imhotep debris falls, represents the occasional exposure of deeper layers which have been subjected to a more complex history. Assuming a size range of tens of micrometres as a typical size for ice grains present on the nucleus of comets (refs 7–9 and this work), the observations of millimetre-sized grains can be explained by the growth of secondary ice crystals from vapour diffusion in ice-rich colder layers and/or by ice grains sintering^{19,20}. The energy necessary to activate these processes is provided by the thermal heat wave penetrating into the subsurface layers and causing ice sublimation; the additional contribution to the total energy balance derived from the amorphous to crystalline phase transition should not be neglected, although it is expected to be moderate in size¹⁹. Ice grain growth can profoundly alter the thermal evolution of the nucleus because it causes a decrease in porosity (densification) and an increase in thermal conductivity. Models show that the ice grain growth and the densification can encompass layers of several metres²⁰, thus affecting the large-scale structural and thermal properties of the nucleus. Moreover, the KOSI experimental results showed that, depending on the dust-to-ice ratio, more than 80% of the sublimating ice was not released through the dust mantle but re-deposited at greater depths²¹. As a consequence, the observed stratification, which is one of the most prominent structural features of comet 67P/Churyumov–Gerasimenko^{11,12}, would be the result of cometary evolution processes rather than a remnant of the nucleus formation scenario.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 August; accepted 28 October 2015.

Published online 13 January 2016.

- Hässig, M. *et al.* Time variability and heterogeneity in the coma of 67P/Churyumov–Gerasimenko. *Science* **347**, <http://dx.doi.org/10.1126/science.aaa0276> (2015).
- Gulkis, S. *et al.* Subsurface properties and early activity of comet 67P/Churyumov–Gerasimenko. *Science* **347**, <http://dx.doi.org/10.1126/science.aaa0709> (2015).
- Mumma, M. J. & Charnley, S. B. The chemical composition of comets—emerging taxonomies and natal heritage. *Annu. Rev. Astron. Astrophys.* **49**, 471–524 (2011).
- Ehrenfreund, P., Charnley, S. B. & Wooden, D. H. in *Comets II* (eds Festou, M. C., Keller, H. U. & Weaver, H. A.) 115–133 (Univ. of Arizona Press, 2004).
- Pommerol, A. *et al.* OSIRIS observations of meter-size exposures of H₂O ice at the surface of 67P/Churyumov–Gerasimenko and interpretation using laboratory experiments. *Astron. Astrophys.* **583**, A25 (2015).
- De Sanctis, M. C. *et al.* The diurnal cycle of water ice on comet 67P/Churyumov–Gerasimenko. *Nature* **525**, 500–503 (2015).
- Capaccioni, F. *et al.* The organic-rich surface of comet 67P/Churyumov–Gerasimenko as seen by VIRTIS/Rosetta. *Science* **347**, <http://dx.doi.org/10.1126/science.aaa0628> (2015).
- Sunshine, J. M. *et al.* Exposed water ice deposits on the surface of comet 9P/Tempel 1. *Science* **311**, 1453–1455 (2006).
- A'Hearn, M. F. *et al.* EPOXI at comet Hartley 2. *Science* **332**, 1396–1400 (2011).
- Sunshine, J. M. *et al.* The distribution of water ice on comet 103P/Hartley 2. *Lunar Planet. Inst. Contrib.* **1667**, 6438 (2012).
- Thomas, N. *et al.* The morphological diversity of comet 67P/Churyumov–Gerasimenko. *Science* **347**, <http://dx.doi.org/10.1126/science.aaa0440> (2015).
- Massironi, M. *et al.* Two independent and primitive envelopes of the bilobate nucleus of comet 67P. *Nature* **526**, 402–405 (2015).
- Sierks, H. *et al.* On the nucleus structure and activity of comet 67P/Churyumov–Gerasimenko. *Science* **347**, <http://dx.doi.org/10.1126/science.aaa1044> (2015).
- Auger, A. T. *et al.* Geomorphology of the Imhotep region on comet 67P/Churyumov–Gerasimenko from OSIRIS observations. *Astron. Astrophys.* **583**, A35 (2015).
- Coradini, A. *et al.* Virts: an imaging spectrometer for the Rosetta mission. *Space Sci. Rev.* **128**, 529–559 (2007).
- Grundy, W. M. & Schmitt, B. The temperature-dependent near-infrared absorption spectrum of hexagonal H₂O ice. *J. Geophys. Res.* **103**, 25809–25822 (1998).

17. Mastrapa, R. M. *et al.* Optical constants of amorphous and crystalline H₂O ice in the near infrared from 1.1 to 2.6 μm . *Icarus* **197**, 307–320 (2008).
18. Sunshine, J. M. *et al.* The distribution of water ice in the interior of comet Tempel 1. *Icarus* **190**, 284–294 (2007).
19. Prialnik, D., Benkhoff, J. & Podolack, M. in *Comets II* (eds Festou, M. C., Keller, H. U. & Weaver, H. A.) 115–133 (Univ. Arizona Press, 2004).
20. Kossacki, K. J., Szutowicz, S. L. & Leliwa-Kopystyński, J. Comet 46P/Wirtanen: evolution of the subsurface layer. *Icarus* **142**, 202–218 (1999).
21. Benkhoff, J., Seidensticker, K. J., Seiferlin, K. & Spohn, T. Energy analysis of porous water ice under space-simulated conditions: results from the KOSI-8 experiment. *Planet. Space Sci.* **43**, 353–361 (1995).

Acknowledgements We thank the following institutions and agencies, which supported this work: Italian Space Agency (ASI, Italy), Centre National d'Etudes Spatiales (CNES, France), Deutsches Zentrum für Luft- und Raumfahrt (DLR, Germany), National Aeronautic and Space Administration (NASA, USA). VIRTIS was built by a consortium from Italy, France and Germany, under the scientific responsibility of the Istituto di Astrofisica e Planetologia Spaziali (IAPS) of INAF, Rome (Italy), which also led the scientific operations. The VIRTIS instrument development for the ESA has been funded and managed by ASI, with contributions from Observatoire de Meudon financed by CNES and from the DLR. The VIRTIS instrument industrial prime contractor was

former Officine Galileo, now Selex ES (Finmeccanica Group) in Campi Bisenzio, Florence, Italy. We also thank the Rosetta Liaison Scientists, the Rosetta Science Ground Segment and the Rosetta Mission Operations Centre for their support in planning the VIRTIS observations. This research has made use of NASA's Astrophysics Data System. This work is dedicated to Angioletta Coradini, conceiver of the VIRTIS instrument.

Author Contributions G.F., M.C.D.S. and F.C. contributed to the data analysis and to the writing of the manuscript. G.F. and F.C. provided calibrated VIRTIS data. A.R. and M.C. provided the spectral fit. F.T. retrieved the temperatures. S.E., S.J., F.T. and C.L. provided geometry information. F.C., G.F., S.E., D.B.-M. and C.L. planned VIRTIS observations with R.N., M.C., A.C. and F.H. implementing telecommands sequences. R.P. and F.H. processed telemetry and data packets. All authors are instrument team members contributing to the discussion of the results.

Author Information The VIRTIS calibrated data will be available through the ESA's Planetary Science Archive (PSA) website by early 2016 (<http://www.rssd.esa.int>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.F. (gianrico.filacchione@iaps.inaf.it).

METHODS

VIRTIS-M. VIRTIS-M is the visible–infrared hyperspectral channel¹⁵ aboard Rosetta. The instrument performs 0.25–5.1- μm spectroscopy by using two separate detectors: a charge-coupled device (CCD) for the 0.25–1- μm range and a HgCdTe for the 1.0–5.1- μm range, respectively. The spectral sampling is equal to 1.8 nm per band and 9.7 nm per band for the two channels, respectively. Both spectral channels share the same telescope equipped with a scanning mirror to build hyperspectral cubes. The instrumental field of view is equal to 3.7° with an instantaneous field of view of 250 μrad . The slit aperture is therefore acquired by 256 spatial samples. The raw data are converted to incident solar flux by applying the calibration pipeline^{22–24} and the corrections derived from in-flight data^{25,26}.

Observation campaigns. During the Philae pre-landing phase, four different nucleus mapping campaigns were carried out by VIRTIS-M: the first in August 2014, during the Medium Term Plan (MTP) mission period (MTP006), when the spacecraft was orbiting between distances of 50–100 km from the nucleus with solar phase angle 25°–40°. In this period many observations were made at constant solar phase (about 30°) with spatial resolution of 12.5–25 m per pixel, including the observation shown in Fig. 2a, in which both debris falls are shown. The second phase occurred in September 2014, during MTP007, when the spacecraft was at a distance of about 30 km with solar phase angle 60°–70°. During this campaign, both debris falls were observed by VIRTIS-M in oblique view with a spatial resolution of about 7.5 m per pixel (Fig. 2b, e). The third campaign (MTP008) was executed in October 2014 when the spacecraft orbited along the nucleus' terminator in circular orbits at distances of 20 km and then 10 km, resulting in observations with a spatial resolution of 5–2.5 m per pixel taken at a solar phase angle of 90°. During the 10-km orbit VIRTIS-M acquired the best view of the waste material at the bottom of BAP1 (see Fig. 2c). Finally, the last campaign was executed from end of October to mid-November 2014 in preparation for the Philae lander release when the spacecraft's trajectory was on an elliptic orbit at distances of between 10 km and 20 km, with VIRTIS-M continuing to monitor the nucleus with spatial resolution between 2.5 m and 5 m per pixel with solar phase angle between 70° to 100°. During this period the visibility of BAP1 continued to be good while BAP2 was no longer visible.

To take into account the local irregular topography we have expanded the range of coordinates over which our spectral analysis is performed to longitude 114°–120° E, latitude 11°–15° N for BAP1 and to longitude 178°–184° E, latitude 2°–12° S for BAP2. Overall, the best viewing of the two BAPs occurs during off-nadir pointing sequences and during Rosetta's terminator orbits, when the emission angle on the exposed walls is smaller. The analysis of VIRTIS geo-referenced images of these areas shows that a total of 7,495 and 7,204 VIRTIS pixels in day-side illumination conditions are located above BAP1 and BAP2, respectively. The corresponding 2- μm water-ice band depth above these points is shown in Fig. 3a–c for the different MTP intervals. Fluctuations are due to the variable water-ice distribution above the local rough topography with different illumination and viewing geometry occurring during Rosetta's orbits around the comet.

Thermal images. For the two better resolved observations of the debris fall regions, thermal images are shown in Fig. 2d–f at the same scale as the corresponding images in reflectance. VIRTIS-M data are used to derive the surface temperature by modelling spectral radiance in the 4.5–5.1- μm range, adopting a Bayesian approach²⁷. The retrieved temperature of a given point on the surface of the nucleus is a function of the local terrain properties (albedo, composition, grain size, roughness, thermal conductivity, volatiles sublimation) and illumination geometry (solar incidence angle, local solar time, time since last shadow occurred). As a consequence of the large obliquity of the spin axis with respect to the orbital plane, the southern hemisphere, where the Imhotep plain is located, is considerably less illuminated than the northern hemisphere until a few months before perihelion passage (which occurred in August 2015). This affects the average temperatures measured on this region, which are much lower than the ones measured on the smaller nucleus lobe and in the north part of the larger one^{2,28}. On the Imhotep plain, the MIRO experiment has measured a maximum diurnal subsurface temperature of about 120 K with both millimetre and submillimetre channels² while modelling VIRTIS-M data a diurnal surface temperature of 180–190 \pm 15 K has been inferred²⁸. Such thermal conditions are favourable for preserving water ice on the surface²⁹. In addition, Fig. 2d–f clearly shows that BAP1 and BAP2 are on average cooler than nearby terrains. In the thermal image of BAP1, VIRTIS-M has measured a temperature lower than 160 \pm 30 K on pixels where water-ice features are more prominent. On BAP2, the temperature is below 180 \pm 15 K. This means that in both regions the water ice is close to the sublimation regime²⁹.

Spectral modelling. A quantitative spectral analysis of the composition of BAP1 and BAP2 has been performed using Hapke's radiative transfer model³⁰ as described in ref. 31. The two regions have been modelled using a complex mixture of two spectral endmembers: crystalline water ice, simulated by using optical constants^{17,32–34} measured at $T = 160$ K between 1 μm and 4 μm , and a 'dark terrain'

unit corresponding to the average spectrum of the comet's surface after the application of photometric correction³⁵. Both endmember spectra are shown in Fig. 3e in the 1.0–2.3 μm spectral range. The analysis has been performed on spectra normalized at 2.3 μm in order to rule out uncertainties on the radiometric and photometric accuracy as well as errors on the local geometry information, owing to unresolved shadows and roughness.

According to Hapke's theory³⁰ the ratio of the bidirectional reflectance $r(i, e, g, \lambda)$ of a semi-infinite medium taken at two wavelengths λ and λ_0 can be expressed as $r(i, e, g, \lambda)/r(i, e, g, \lambda_0)$, where:

$$r(i, e, g, \lambda) = \frac{w(\lambda)K}{4\pi} \frac{\mu_{0e}}{\mu_{0e} + \mu_e} [B_{SH}(g)p(g, \lambda) + H[w(\lambda), \mu_{0e}/K]H[w(\lambda), \mu_e/K] - 1]S(i, e, g, \theta)B_{CB}(g, \lambda) \quad (1)$$

and

$$r(i, e, g, \lambda_0) = \frac{w(\lambda_0)K}{4\pi} \frac{\mu_{0e}}{\mu_{0e} + \mu_e} [B_{SH}(g)p(g, \lambda_0) + H[w(\lambda_0), \mu_{0e}/K]H[w(\lambda_0), \mu_e/K] - 1]S(i, e, g, \theta)B_{CB}(g, \lambda_0) \quad (2)$$

Here i , e and g are the incidence, emission and phase angle, respectively; w is the single scattering albedo; K is the porosity of the medium; $p(g, \lambda)$ is the single particle phase function; μ_{0e} and μ_e are the effective cosines of the incidence and emission angles, differing from μ_0 and μ because of the effect of surface roughness; $H(w, \mu_e/K)$ is the Chandrasekhar function describing the multiple scattering components; $B_{SH}(g)$ is the shadow hiding opposition effect term; $B_{CB}(g, \lambda)$ the coherent back-scattering opposition effect; $S(i, e, g, \theta)$ is the shadow function modelling large-scale roughness and θ is the average surface slope.

Taking the ratio $r(i, e, g, \lambda)/r(i, e, g, \lambda_0)$, all the terms that do not depend on wavelength will cancel each other out. In particular, the porosity K and the roughness embedded in the $S(i, e, g, \theta)$ shadow function are removed. Moreover, during the VIRTIS-M observations described in this paper the solar phase was fixed at $g = 95^\circ$. As a consequence the shadow hiding effect is negligible and this implies that the terms B_{SH} and B_{CB} are equal to 1.

In summary, the ratio $r(i, e, g, \lambda)/r(i, e, g, \lambda_0)$ will be simplified as follows:

$$\frac{r(i, e, g, \lambda)}{r(i, e, g, \lambda_0)} = \frac{w(\lambda)}{w(\lambda_0)} \left[\frac{p(g, \lambda) + H[w(\lambda), \mu_{0e}/K]H[w(\lambda), \mu_e/K] - 1}{p(g, \lambda_0) + H[w(\lambda_0), \mu_{0e}/K]H[w(\lambda_0), \mu_e/K] - 1} \right] \quad (3)$$

In addition, for the dark terrain component, which makes up the majority by far of the material present on the nucleus surface, the contribution of the porosity and of the roughness to the multiple scattering term is a second-order effect. Thus we can assume $\mu_{0e}/K = \mu_0$.

Equation (2) represents the formulation to be used to derive the normalized reflectance of a single material but it can be straightforwardly extended to the case of a linear combination of reflectances, maintaining the simplifications described above. In this relation the parameters to fit are the phase function $p(g)$ and the single scattering albedo w , which can be modelled using the optical constants and the grain size. The phase function $p(g)$ derived by ref. 35 is used in the spectral modelling.

We are dealing with two endmembers and we have used areal and intimate mixing modalities in the simulations: in the areal mixing the surface is modelled as patches of pure water ice and dark terrain, with each photon scattered within one patch. In the intimate mixing model the particles of the two endmember materials are in contact with each other and both are involved in the scattering of a single photon. In intimate mixing the single scattering albedo of the mixture is the weighted average, through their abundance P , of the dark terrain and water-ice single scattering albedos. Thus the $w(\lambda)$ is given by $w(\lambda) = w(\lambda)_{DT}P_{DT} + w(\lambda)_{WI}P_{WI}$. However, in the areal mixing (linear mixing), the two components are treated independently. For the dark terrain we have used the $w(\lambda)$ value retrieved from the photometric correction³⁵, while to describe the water ice contribution we explicitly calculated the $w(\lambda)$ from the optical constants and from the grain size, which represent the parameter to be fitted. Furthermore, after extensive tests, we have been able to fix the value of the phase function $p(g = 95^\circ) = 0.5$. With the approach described above, the model's free parameters are the percentage and grain size of the water ice in the intimate mixed phase, and the percentage and grain size of the water ice in the areal mixed phase.

Before fitting, the observed spectra are corrected for spikes and for instrumental artefacts. The best-fitting result, an example of which is shown in Fig. 4a, is obtained by applying the Levenberg–Marquardt method for nonlinear least-squares multiple regression. In this specific case, the best matching spectrum corresponds to an areal mixing where $a_{H_2O} = 1.2\%$ of the pixel area is modelled with large grains (size distribution peaked around $d_{H_2O} = 2$ mm in diameter) of pure

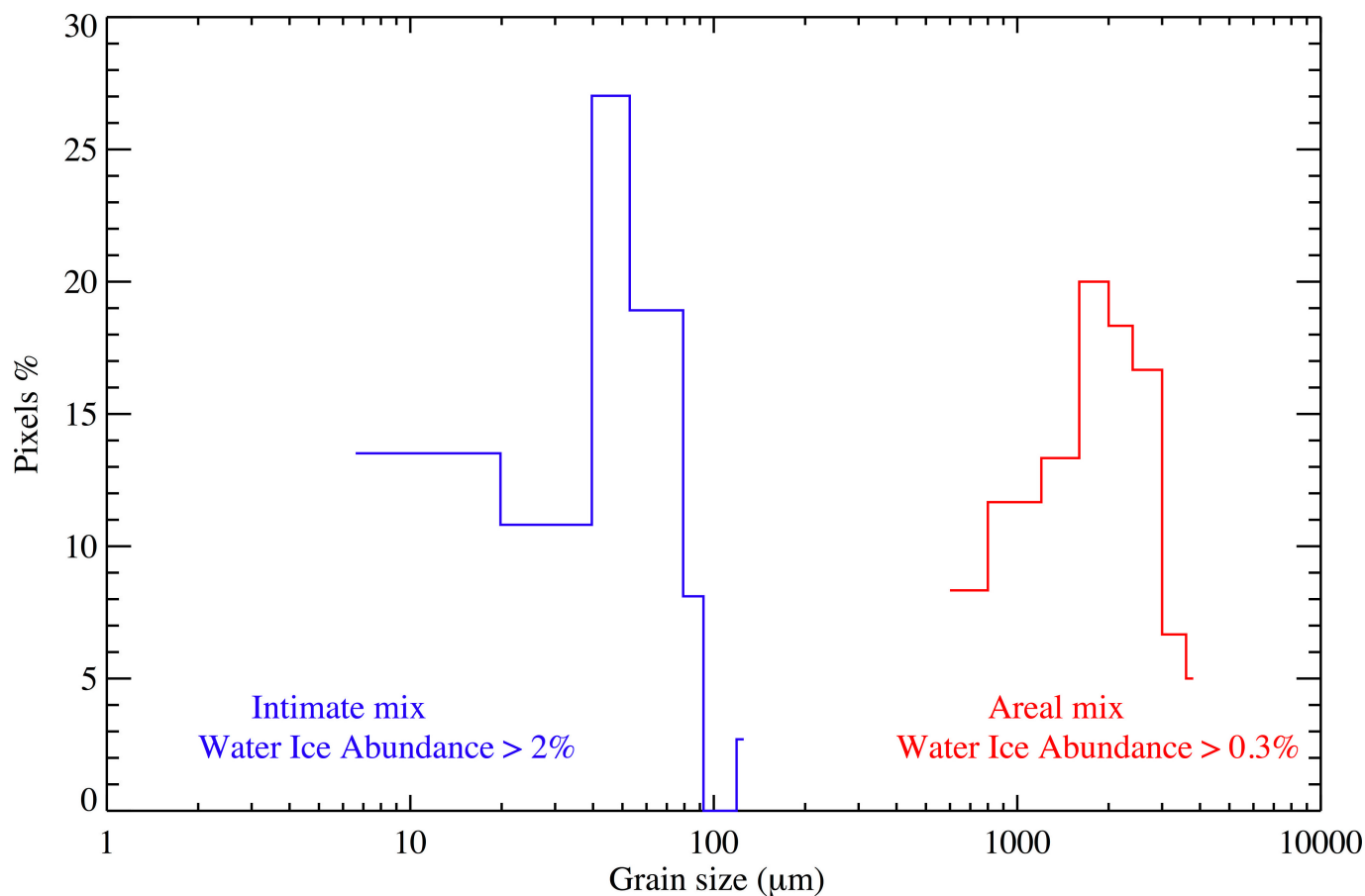
water ice while the remaining 98.8% of the pixel area is composed of an intimate mixture with $a_{\text{H}_2\text{O}} = 3.4\%$ of water ice (size distribution peaked at $d_{\text{H}_2\text{O}} = 56 \mu\text{m}$ in diameter) and the remaining 95.4% of dark material. The distribution of the areal and intimate grain sizes appears to be monodispersed (Extended Data Fig. 1) and grouped into two very different families: small grains of 50–60 μm diameter size in the intimate case and large millimetre-size grains in the areal case. When we say that our fit requires particles of average sizes of 2 mm, this means that, from a photon point of view, the medium is essentially continuous at scales of 2 mm. This implies that any discontinuities (for example, voids) within a volume of 2 mm radius should be smaller than the wavelength.

NAVCAM images references. The Rosetta NAVCAM context images of the two debris fall shown in Fig. 1 are: ROS_CAM1_20140823T184003, taken on 2014-08-23T18:40 (corresponding to 23 August 2014 at 18:40 Universal Standard Time) from a distance of 64.8 km (Fig. 1a). BAP1 on the Ash–Khepry–Imhotep boundary (at longitude 117° E, latitude 13° N) is indicated by the white arrow. BAP2 on the hillside (at longitude 180°–182° E, latitude 4°–10° S) is indicated by the yellow arrow; ROS_CAM1_20140914T121933, taken on 2014-09-14T12:19 from a distance of 30.7 km (Fig. 1b). The image shows the waste material and debris accumulated at the bottom of the landslide eroding the side of the elevated circular structure. BAP1 is not visible in the viewing geometry of this nadir-pointing image; ROS_CAM1_20140920T060854, taken on 2014-09-20T06:08 from a distance of 28.2 km (Fig. 1c). BAP1 is caused by the presence of water ice, clearly visible on the oblique view of the elevated structure wall; ROS_CAM1_20140915T021834, taken on 2014-09-15T02:18 from a distance of 29.9 km (Fig. 1d). This image shows a viewing geometry identifying BAP2 similar to Fig. 1c. Note a secondary smaller bright layer on the right of the primary.

VIRTIS-M hyperspectral cube references. The Rosetta VIRTIS-M observations of the two BAPs shown in Fig. 2 are as follows: I1_00367585390, taken on 2014-08-25T11:04 with an integration time of 3 s per line and a spatial resolution of 12.5 m per pixel (Fig. 2a); I1_00369364114, taken on 2014-09-15T01:09 with an integration time of 3 s per line and a spatial resolution of 7.5 m per pixel (Fig. 2b); I1_00371998843, taken on 2014-10-15T13:01 with an integration time of 3 s per line and a spatial resolution of 2.5 m per pixel (Fig. 2c, d); I1_00369369214, taken

on 2014-09-15T02:34 with an integration time of 3 s per line and a spatial resolution of 7.5 m per pixel (Fig. 2e, f).

22. Filacchione, G. Calibrations a terra e prestazioni in volo di spettrometri ad immagine nel visibile e nel vicino infrarosso per l'esplorazione planetaria. PhD dissertation, Univ. Studi di Napoli Federico II (2006); available at http://www.fedoa.unina.it/1462/1/Filacchione_Ingegneria_Aerospaziale_Navale_e_della_Qualita.pdf.
23. Ammannito, E. *et al.* On-ground characterization of Rosetta/VIRTIS-M. I. Spectral and geometrical calibrations. *Rev. Sci. Instrum.* **77**, 093109 (2006).
24. Filacchione, G. *et al.* On-ground characterization of Rosetta/VIRTIS-M. II. Spatial and radiometric calibrations. *Rev. Sci. Instrum.* **77**, 103106 (2006).
25. Migliorini, A. *et al.* Comparative analysis of airglow emissions in terrestrial planets, observed with VIRTIS-M instruments on board Rosetta and Venus Express. *Icarus* **226**, 1115–1127 (2013).
26. Raponi, A. Spectrophotometric analysis of cometary nuclei from in situ observations. PhD thesis, Univ. degli studi di Roma Tor Vergata (2014); preprint at <http://arxiv.org/abs/1503.08172>.
27. Tosi, F. *et al.* Thermal measurements of dark and bright surface features on Vesta as derived from Dawn/VIR. *Icarus* **240**, 36–57 (2014).
28. Tosi, F. *et al.* Thermal maps and properties of comet 67P as derived from Rosetta/VIRTIS data. *Lunar Planet. Sci. Conf. XXXVI*, 2156 (LPI contribution no. 1832, 2015).
29. Fray, N. & Schmitt, B. Sublimation of ices of astrophysical interest: a bibliographic review. *Planet. Space Sci.* **57**, 2053–2080 (2009).
30. Hapke, B. *Theory of Reflectance and Emittance Spectroscopy* (Cambridge Univ. Press, 2012).
31. Ciarniello, M. *et al.* Hapke modeling of Rhea surface properties through Cassini-VIMS spectra. *Icarus* **214**, 541–555 (2011).
32. Ciarniello, M. *et al.* Photometric properties of comet 67P/Churyumov–Gerasimenko from VIRTIS-M onboard Rosetta. *Astron. Astrophys.* **583**, A31 (2015).
33. Warren, S. G. Optical constants of ice from the ultraviolet to the microwave. *Appl. Opt.* **23**, 1206 (1984).
34. Mastrapa, R. M. *et al.* Optical constants of amorphous and crystalline H₂O-ice: 2.5–22 μm (4000–455 cm^{-1}) optical constants of H₂O-ice. *Astrophys. J.* **701**, 1347–1356 (2009).
35. Clark, R. N. *et al.* The surface composition of Iapetus: mapping results from Cassini VIMS. *Icarus* **218**, 831–860 (2012).



Extended Data Figure 1 | Water-ice grain size distribution derived from BAP1. Grains are present in two monodispersed distributions with maxima at 56 μm and at 2 mm, corresponding to the intimate and areal

mixing classes, respectively. The histogram is computed by selecting only pixels showing a water-ice abundance greater than 2% for the intimate mixing class and greater than 0.3% for areal mixing.

An improved limit on the charge of antihydrogen from stochastic acceleration

M. Ahmadi¹, M. Baquero-Ruiz^{2,3}, W. Bertsche^{4,5}, E. Butler^{6,7}, A. Capra⁸, C. Carruth², C. L. Cesar⁹, M. Charlton¹⁰, A. E. Charman², S. Eriksson¹⁰, L. T. Evans², N. Evetts¹¹, J. Fajans², T. Friesen¹², M. C. Fujiwara¹³, D. R. Gill¹³, A. Gutierrez¹¹, J. S. Hangst¹², W. N. Hardy¹¹, M. E. Hayden¹⁴, C. A. Isaac¹⁰, A. Ishida⁷, S. A. Jones¹⁰, S. Jonsell¹⁵, L. Kurchaninov¹³, N. Madsen¹⁰, D. Maxwell¹⁰, J. T. K. McKenna¹³, S. Menary⁸, J. M. Michan¹³, T. Momose¹⁶, J. J. Munich¹⁴, P. Nolan¹, K. Olchanski¹³, A. Olin^{13,17}, A. Povilus², P. Pusa¹, C. Ø. Rasmussen¹², F. Robicheaux¹⁸, R. L. Sacramento⁹, M. Sameed¹⁰, E. Sarid¹⁹, D. M. Silveira⁹, C. So², T. D. Tharp¹², R. I. Thompson²⁰, D. P. van der Werf¹⁰, J. S. Wurtele^{2,21} & A. I. Zhmoginov²

Antimatter continues to intrigue physicists because of its apparent absence in the observable Universe. Current theory requires that matter and antimatter appeared in equal quantities after the Big Bang, but the Standard Model of particle physics offers no quantitative explanation for the apparent disappearance of half the Universe. It has recently become possible to study trapped atoms^{1–4} of antihydrogen to search for possible, as yet unobserved, differences in the physical behaviour of matter and antimatter. Here we consider the charge neutrality of the antihydrogen atom. By applying stochastic acceleration to trapped antihydrogen atoms, we determine an experimental bound on the antihydrogen charge, Q_e , of $|Q| < 0.71$ parts per billion (one standard deviation), in which e is the elementary charge. This bound is a factor of 20 less than that determined from the best previous measurement⁵ of the antihydrogen charge. The electrical charge of atoms and molecules of normal matter is known⁶ to be no greater than about $10^{-21}e$ for a diverse range of species including H_2 , He and SF_6 . Charge-parity-time symmetry and quantum anomaly cancellation⁷ demand that the charge of antihydrogen be similarly small. Thus, our measurement constitutes an improved limit and a test of fundamental aspects of the Standard Model. If we assume charge superposition and use the best measured value of the antiproton charge⁸, then we can place a new limit on the positron charge anomaly (the relative difference between the positron and elementary charge) of about one part per billion (one standard deviation), a 25-fold reduction compared to the current best measurement^{8,9}.

The charge of the antihydrogen atom can be inferred from the charge anomalies of the antiproton and positron. The charge anomaly of the antiproton is known from spectroscopic measurements^{8,10} on antiprotonic helium, $\bar{p}He^+$, and is $|q_{\bar{p}} - e|/e < 0.7$ parts per billion (p.p.b.) with a confidence level of 90%, where $q_{\bar{p}}$ is the antiproton charge. The charge anomaly of the positron^{8,9,11} is less well known: $|q_{e^+} - e|/e < 25$ p.p.b., where q_{e^+} is the positron charge (no confidence level given, but assumed here to be 1σ), determined by measurements of the positron cyclotron frequency and the positronium Rydberg constant. Assuming charge superposition, we can therefore combine these two anomalies to infer an indirect experimental bound of $|Q| \lesssim 25$ p.p.b. on the charge Q_e of antihydrogen. This bound is much looser than the

direct experimental bounds for normal-matter atoms; however, the methods used to set these normal-matter bounds generally use macroscopic quantities of atoms and so are unsuitable for studies of antihydrogen, of which only approximately 1,000 antiatoms have been trapped and detected^{1–4}.

The earliest direct experimental limit on the charge of antihydrogen, $|Q| < 0.02$, came from the observation¹² that Lorentz forces were insufficient to deflect an energetic (momentum of $1.94 \text{ GeV } c^{-1}$) beam¹³ of antihydrogen atoms away from a detector. A much more precise bound, $Q = (-13 \pm 11 \pm 4)$ p.p.b. (1σ confidence level), was recently determined⁵ by searching for a static-electric-field-induced shift in the average position of antiatoms confined in the ALPHA-I antihydrogen trap at CERN. This shift was measured by recording the position along the electric-field axis as the antiatoms were released from the trap by turning off the trapping magnetic fields.

The precision of this static shift measurement was limited by the magnitude of the applied electric fields, and by statistical and systematic issues related to accurately measuring small deflections. The sensitivity of the charge measurement can be enhanced by using stochastic (that is, randomly time-varying) electric fields^{5,14,15}. Under the influence of such fields, a putatively charged, trapped antihydrogen atom would experience small, stochastic energy kicks. On average, the antiatom would gradually gain enough energy from these kicks to escape the shallow potential well in which it was trapped. This process can be modelled as a random walk similar to Fermi acceleration¹⁶. By searching for the loss of antiatoms caused by these stochastic electric fields, we set a new limit on $|Q|$.

Our measurement was performed on antihydrogen atoms confined in the ALPHA-2 trap at CERN. This trap uses a minimum-magnetic-field (minimum- B) configuration to trap the antiatoms. Schematics of the trap and fields are shown in Fig. 1a and b, respectively. The magnetic minimum is formed by mirror and octupole magnetic coils. Antihydrogen atoms whose magnetic moments are oriented antiparallel to the trap magnetic fields are attracted to the magnetic minimum at the trap centre; for our fields, which are of the order of 1 T, the depth of the antihydrogen potential well is $E_{\text{well}} = 0.54 \text{ K}$, where E_{well} is measured in temperature units.

The general methods by which we trap antihydrogen are described elsewhere^{1–3,17}, but briefly, antihydrogen atoms are created by first trapping separate cold plasmas containing several million positrons

¹Department of Physics, University of Liverpool, Liverpool L69 7ZE, UK. ²Department of Physics, University of California at Berkeley, Berkeley, California 94720-7300, USA. ³Centre de Recherches en Physique des Plasmas (CRPP), École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. ⁴School of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK. ⁵Cockcroft Institute, Sci-Tech Daresbury, Warrington WA4 4AD, UK. ⁶Centre for Cold Matter, Blackett Laboratory, Imperial College London, Prince Consort Road, London SW7 2AZ, UK. ⁷Physics Department, European Organisation for Nuclear Research (CERN), CH-1211 Geneva 23, Switzerland. ⁸Department of Physics and Astronomy, York University, Toronto, Ontario M3J 1P3, Canada. ⁹Instituto de Física, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21941-972, Brazil. ¹⁰Department of Physics, Swansea University, Swansea SA2 8PP, UK. ¹¹Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. ¹²Department of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark. ¹³TRIUMF, 4004 Wesbrook Mall, Vancouver, British Columbia V6T 2A3, Canada. ¹⁴Department of Physics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada. ¹⁵Department of Physics, Stockholm University, SE-10691, Stockholm, Sweden. ¹⁶Department of Chemistry, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada. ¹⁷Department of Physics and Astronomy, University of Victoria, Victoria, British Columbia V8P 5C2, Canada. ¹⁸Department of Physics and Astronomy, Purdue University, West Lafayette, Indiana 47907, USA. ¹⁹Soreq Nuclear Research Center, Yavne, 81800, Israel. ²⁰Department of Physics and Astronomy, University of Calgary, Calgary, Alberta T2N 1N4, Canada. ²¹ATAP, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

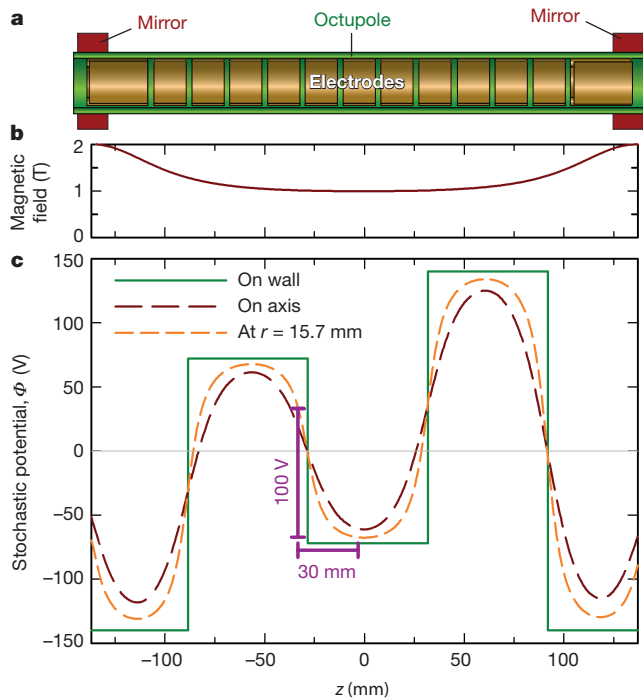


Figure 1 | Experiment schematic. **a**, Electrode structure of the ALPHA-2 trap, showing the approximate positions of the mirror and octupole coils. The particle detector (not shown) surrounds the octupole. (The vertical scale has been expanded for clarity.) **b**, Magnetic field on the axis (z) of the atom trap showing the axial magnetic potential well induced by the mirror fields. **c**, A snapshot of the applied stochastic electric-field potential Φ at several radii r as a function of z . ‘On wall’ indicates the wall of the Penning trap; ‘on axis’ indicates the axis of the trap ($r=0$). The purple brackets indicate a typical travel distance and potential change experienced by an antiatom with 0.1 K of axial energy in 1 ms.

and several tens of thousands of antiprotons. The positrons come from a radioactive sodium source¹⁸, and the antiprotons are generated by CERN’s Antiproton Decelerator (AD). Antiatoms are formed by three-body recombination when these two plasmas are mixed using autoresonance^{19,20} in the trap interior. Approximately 0.01% of the roughly ten thousand atoms created in each 1-s-long mixing interval are sufficiently cold to be trapped. The antiatoms are detected by releasing them from the trap by turning off the trap magnetic fields. A three-layer silicon vertex detector (SVD) surrounding the trap is used to detect, locate and time the resulting antiatom annihilations²¹. The detector and its associated multivariate analysis^{22–24} (MVA) software can detect approximately 82% of the antiatoms that annihilate in the trap.

To generate the stochastic electric field necessary for our measurement, we impose the electric potential shown in Fig. 1c on the trap by biasing the trap electrodes. We make this potential, and the associated electric field, vary stochastically by repeatedly changing the bias on the electrodes from that shown in Fig. 1c to its negation. We use $N=84,900$ transitions, which, altogether, last 114.9 s in each experimental cycle (see Fig. 2). The interval between transitions averages 1 ms, but is randomized for reasons that are discussed later.

Each of the potential transitions would non-adiabatically perturb the kinetic energy of a charged antiatom; as noted above, these kicks correspond, approximately, to a random walk in energy. To model this process, we need to estimate the average potential change subsequent to each of the transitions. Measurements using our first trap, ALPHA-I³, showed that the energy of the trapped antiatoms is roughly distributed^{12,5,25,26} as a high-temperature, three-dimensional (3D) Maxwellian distribution truncated near E_{well} . The typical energy of the component of motion along the trap axis is therefore about 0.1 K, which corresponds to a travel distance of about 30 mm in 1 ms. From the purple

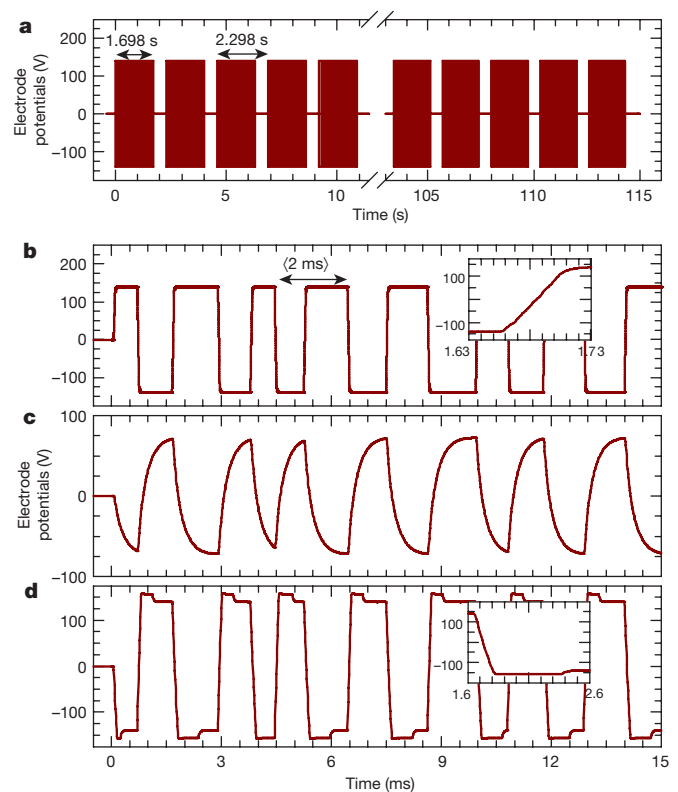


Figure 2 | Stochastic drive potentials. **a**, Time structure of the stochastic drive potential repetitions. The potential is non-zero within the first 1.698 s of each repetition. The fifty repetitions are separated by 2.298 s. **b–d**, Snapshots of the typical time histories of the measured potentials applied to different electrodes; insets illustrate details of the transitions. Note that the time intervals between transitions vary stochastically.

brackets shown in Fig. 1c, we see that an antiatom possessing this axial kinetic energy would experience a potential change of $\Delta\Phi \approx 100$ V in 1 ms, although there are substantial uncertainties in this estimate due to variations in the potential over the volume of the trap, the distribution of kinetic energies and the potential changes arising from radial motion. A simple energy-diffusion model predicts that an antiatom with a putative charge Qe would gain energy of about $|Q|e\Delta\Phi\sqrt{N}$. Consequently, the antiatom would gain roughly enough energy to escape from a trap of depth E_{well} if

$$|Q| \gtrsim \frac{E_{\text{well}}}{e\Delta\Phi\sqrt{N}} \quad (1)$$

Evaluating this equation using our experimental parameters yields $|Q| \gtrsim 1.6$ p.p.b. However, equation (1) assumes that the antiatom starts with zero initial kinetic energy. Prior measurements indicate that the average antiatom starts with more than half the energy it needs to escape; thus, the bound on $|Q|$ is probably lower.

Stochastic acceleration requires an element of randomness in the relation between the frequency of the driving-potential inversions and the frequencies of the antiatom orbital oscillations. Trajectory simulations (see, for example, figure 6 in ref. 15) demonstrate that the orbits in the unforced system are generally not regular and do not have a fixed frequency. In many cases, these variations in orbital frequency would make even a strictly periodic driving potential act stochastically. However, some orbits have long-lasting periods of regular motion, and antiatoms in these orbits might not be accelerated by the driving potential. As proposed in ref. 15, we introduce sufficient stochasticity into the system by a random modulation, with a uniform distribution and a standard deviation of 0.2 ms, of the time interval between potential transitions. (This variability is visible in Fig. 2b–d.) Simulations^{14,15} show that it is not necessary

Table 1 | Results of the stochastic and null trials

	Number of trials	Observed antiatoms
Stochastic trials	10	12
Null trials	10	12

or beneficial to randomize the voltage levels between which the electrode potentials switch.

To bound $|Q|$, we alternated between two experimental protocols, each with four phases. The first protocol involved stochastic trials, in which we (1) trapped antihydrogen, (2) subjected any trapped antihydrogen to the stochastic potential variations described above, (3) allowed any antiatoms remaining after the stochastic phase to escape by turning the confining magnetic fields off, and then (4) counted the number of antiatoms that thus escaped using the silicon vertex detector. The second protocol involved null trials, which were performed in strict alternation with the stochastic trials, and were identical to them in all respects, except that the potentials were set to zero during phase (2). Thus, in both types of trials we attempted to hold the antiatoms for the same time (114.9 s) after phase (1). Assuming that the underlying initial antihydrogen trapping rate and all other systematic effects are the same in the stochastic and null trials, a statistically significant difference between the number of antiatoms detected after these trials would imply the existence of a non-zero Q being acted upon by the oscillating electric fields of the stochastic trials. (As discussed in Methods, known polarization and polarizability effects are negligible in the current experiment.)

The results of executing both protocols ten times are presented in Table 1. We observed no differences in the number of antiatoms released in phase (3) between the two protocols, which suggests that the charge—if any—of antihydrogen is below that required to escape from the trap, $|Q| \lesssim 1.6$ p.p.b. (based on the analytic estimate given above). This rough estimate can be made more robust and precise by simulating the dynamics of the antiatoms and using a proper statistical analysis. (Events detected during phase (2) are compiled in Extended Data Table 1 and discussed in Methods, and are compatible with cosmic rays.)

We use simulations^{1,2,5,15,25,26} that model the experiment closely, including detailed models of the trap fields and stochastic timing, and realistic initial conditions validated by previous experiments. The simulations begin by mimicking the initial mixing of the two plasmas and subsequent relaxation periods. We model the null trials by propagating the antiatoms in the electric-field-free trap for a further 114.9 s. We model the stochastic trials by assigning the antiatoms a putative Q , and propagating them in the presence of the stochastic electric fields for this same 114.9 s. For each value of Q studied, we launch 1,000 antiatoms.

For each simulated Q , we determine the survival probability s by calculating the ratio of the number of antiatoms that survive the stochastic simulations to the number that survive the null simulations; the results are shown in Fig. 3. Because the stochastic drive reverses direction every half-cycle, it is unbiased on average; consequently, s depends on the magnitude of the charge $|Q|$, but not on its sign.

Although it is clear from Fig. 3 that the bound on $|Q|$ set by our experiment will be in the neighbourhood of 1 p.p.b., it is not immediately clear exactly what s and, hence, what $|Q|$ to assign from this figure and the data in Table 1. We might begin by setting a threshold of $s = 0.5$; however with a per-trial antiatom observation rate of $r \approx 24/20 = 1.2$, the probability that 12 (what we observed) or more antiatoms would have survived in 10 trials with $s = 0.5$ is only 0.02 assuming Poissonian statistics. Clearly, our data will support setting the threshold at some larger value of s . Furthermore, our limited data set only provides an estimate for r . This suggests that a Bayesian methodology, in which r is treated as a nuisance parameter, would be useful for finding a more precise limit on $|Q|$.

In Methods we present a simple Bayesian analysis that suggests that $s > 0.79$ and that the corresponding bound on $|Q|$, found by inverting the curve in Fig. 3, is $|Q| < 0.59$ p.p.b. However, this analysis ignores

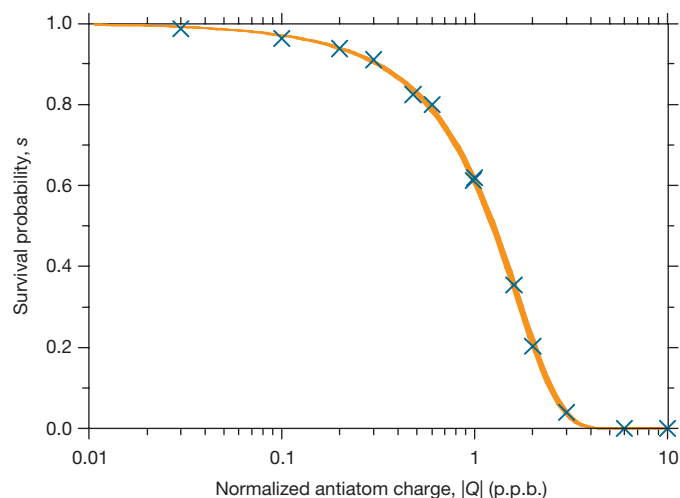


Figure 3 | Survival probability. Simulated survival probability s as a function of $|Q|$ for the stochastic trials. The blue crosses are the number of antiatoms surviving at the given $|Q|$ value divided by the number of antiatoms surviving at $Q = 0$ (the null simulation). The orange band of varying thickness is the 1σ confidence region from a Bayesian fit to this simulation data; see the Supplementary Information for further explanation.

systematic uncertainties; fortunately, our experiment is relatively insensitive to these uncertainties. For example, the analytic expression given in equation (1) shows that uncertainties in Φ (or in the magnitude of the electric field $|E|$) affect our bound on $|Q|$ linearly. Because these fields are known to within approximately 1%, uncertainties in Φ have a negligible effect compared to that of our statistical uncertainties. Similarly, uncertainties in the magnitude of the magnetic field $|B|$ contribute to uncertainties in $|Q|$ through its linear dependency on E_{well} . Because E_{well} depends on the mirror and octupole fields approximately linearly, and because both are known to within approximately 1%, these effects are unimportant. Details of the magnetic field structure beyond the maximum and minimum are similarly unimportant because they do not affect the trap depth. Simulations show that s is not greatly influenced by changes in the volume from which the antiatoms are initially launched.

All of the data were taken over adjacent half-shifts, separated by 16 h, at CERN's AD facility. We are not aware of any mechanism that could have substantially altered the observation rate r in synchronization with the strict alternation of the stochastic and null trials. Likewise, this alternation makes the measurement robust against failures of the detector system. Nonetheless, we performed extensive tests²⁷ of the detector performance to confirm that there were no substantial failures or performance degradations (see Extended Data Fig. 1). The expected number of cosmic rays incorrectly identified as antiatoms is 0.07 for each of the two sets of ten trials; this has a negligible effect on our bound on $|Q|$ (see Supplementary Information).

Simulations do reveal a small dependence of s on the initial energy distribution of the antiatoms. Two distributions that bound our preferred, truncated Maxwellian distribution were studied for our previous apparatus ALPHA-I. Although we have good evidence to rule out these bounding distributions^{5,25,26}, their impact on the measurement reported here is discussed in Supplementary Information. With an additional, even smaller correction arising from the statistical error in the Monte Carlo simulation, our bound degrades to $|Q| < 0.71$ p.p.b. at a 68.3% (1σ) confidence level. This bound represents a 20-fold improvement on the best previous bound⁵.

Given that the charge anomaly of the antiproton^{8,10} is bound at below 0.7 p.p.b., our result, combined with the assumption of charge superposition, limits the charge anomaly of the positron to 1 p.p.b. (1σ), a 25-fold improvement on the best previous bound^{8,9}. By modestly increasing the potentials, holding the antiatoms for longer times and cooling the antiatoms²⁸, we expect (see the scalings in equation (1)) to

be able to improve the antihydrogen charge bound by several orders of magnitude¹⁵; these realizable improvements exceed those that could be obtained by the previously used static-electric-field shift methodology⁵.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 September; accepted 17 November 2015.

- Andresen, G. B. *et al.* Trapped antihydrogen. *Nature* **468**, 673–676 (2010).
- Andresen, G. B. *et al.* Confinement of antihydrogen for 1,000 seconds. *Nature Phys.* **7**, 558–564 (2011).
- Amole, C. *et al.* The ALPHA antihydrogen trapping apparatus. *Nucl. Instrum. Methods Phys. Res. A* **735**, 319–340 (2014).
- Gabrielse, G. *et al.* Trapped antihydrogen in its ground state. *Phys. Rev. Lett.* **108**, 113002 (2012).
- Amole, C. *et al.* Experimental limit on the charge of antihydrogen. *Nature Commun.* **5**, 3955 (2014).
- Bressi, G. *et al.* Testing the neutrality of matter by acoustic means in a spherical resonator. *Phys. Rev. A* **83**, 052101 (2011).
- Quigg, C. *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions* (Westview Press, 1997).
- Olive, K. A. *et al.* Review of particle physics. *Chinese Phys. C* **38**, 090001 (2014).
- Fee, M. S. *et al.* Measurement of the positronium 1^3S_1 – 2^3S_1 interval by continuous-wave two-photon excitation. *Phys. Rev. A* **48**, 192–219 (1993).
- Hori, M. *et al.* Two-photon laser spectroscopy of antiprotonic helium and the antiproton-to-electron mass ratio. *Nature* **475**, 484–488 (2011).
- Hughes, R. J. & Deutch, B. I. Electric charges of positrons and antiprotons. *Phys. Rev. Lett.* **69**, 578–581 (1992).
- Greenland, P. T. Antimatter. *Contemp. Phys.* **38**, 181–203 (1997).
- Baur, G. *et al.* Production of antihydrogen. *Phys. Lett. B* **368**, 251–258 (1996).
- Baquero-Ruiz, M. *Studies on the Neutrality of Antihydrogen*. PhD thesis, University of California, Berkeley (2013).
- Baquero-Ruiz, M. *et al.* Using stochastic acceleration to place experimental limits on the charge of antihydrogen. *New J. Phys.* **16**, 083013 (2014).
- Fermi, E. On the origin of the cosmic radiation. *Phys. Rev.* **75**, 1169–1174 (1949).
- Andresen, G. B. *et al.* Search for trapped antihydrogen. *Phys. Lett. B* **695**, 95–104 (2011).
- Murphy, T. & Surko, C. Positron trapping in an electrostatic well by inelastic collisions with nitrogen molecules. *Phys. Rev. A* **46**, 5696–5705 (1992).
- Fajans, J. & Friedland, L. Autoresonant (nonstationary) excitation of a pendulum, Plutinos, plasmas and other nonlinear oscillators. *Am. J. Phys.* **69**, 1096–1102 (2001).
- Andresen, G. B. *et al.* Autoresonant excitation of antiproton plasmas. *Phys. Rev. Lett.* **106**, 025002 (2011).
- Amole, C. *et al.* Silicon vertex detector upgrade in the ALPHA experiment. *Nucl. Instrum. Methods Phys. Res. A* **732**, 134–136 (2013).
- Narsky, I. StatPatternRecognition: a C++ package for statistical analysis of high energy physics data. Preprint at <http://arxiv.org/abs/physics/0507143> (2005).


- Narsky, I. Optimization of signal significance by bagging decision trees. Preprint at <http://arxiv.org/abs/physics/0507157> (2005).
- Amole, C. *et al.* Resonant quantum transitions in trapped antihydrogen atoms. *Nature* **483**, 439–443 (2012).
- Amole, C. *et al.* Discriminating between antihydrogen and mirror-trapped antiprotons in a minimum-B trap. *New J. Phys.* **14**, 015010 (2012).
- Amole, C. *et al.* Description and first application of a new technique to measure the gravitational mass of antihydrogen. *Nature Commun.* **4**, 1785 (2013).
- Capra, A. *Testing CPT and antigravity with trapped antihydrogen at ALPHA*. PhD thesis, York University (2015).
- Donnan, P. H., Fujiwara, M. C. & Robicheaux, F. A proposal for laser cooling antihydrogen atoms. *J. Phys. B* **46**, 025302 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements All authors except A.E.C. are members of the ALPHA Collaboration. This work was supported by: CNPq, FINEP-RENAFAE (Brazil); NSERC, NRC/TRIUMF, AITF, FQRNT (Canada); FNU, Carlsberg Foundation (Denmark); JSPS Postdoctoral Fellowships for Research Abroad (Japan); ISF (Israel); STFC, EPSRC, the Royal Society and the Leverhulme Trust (UK); DOE, NSF (USA); and VR (Sweden). We thank C. Marshall, J. Strachan and their teams at TRIUMF and Daresbury Laboratory for their contributions to the design and manufacturing of the ALPHA-2 apparatus, and are grateful for the efforts of the CERN AD team, without which these experiments could not have taken place. We thank A. Ballarino, P. Denis and A. Perin for their help in providing and implementing the high-temperature superconductor current leads for the ALPHA-2 magnets.

Author Contributions This analysis was based on data collected on the ALPHA-2 antihydrogen trapping apparatus constructed by the ALPHA Collaboration using methods developed by the entire collaboration. Interest in a charge measurement was first expressed by M.C.F.; the stochastic acceleration methodology used here was proposed by J.F. and further developed by M.B.-R., W.B., A.E.C., L.T.E., F.R., J.S.W. and A.I.Z. The detector systematic studies were undertaken by A.C., D.R.G., M.C.F., J.T.K.McK., S.M., A.O. and P.P. This article was written by A.E.C., J.F., L.T.E. and J.S.W., with extensive help from J.S.H. and M.E.H., and then improved and approved by all authors. This article constitutes part of the PhD work of M.B.-R. and A.C.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to ALPHA Collaboration (alpha-contact@cern.ch), or to J.F. (joel@physics.berkeley.edu) or J.S.W. (wurtele@berkeley.edu).

 This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>.

METHODS

Stochastic potentials. The stochastic potentials are applied by biasing the electrically isolated trap electrodes shown in Fig. 1a. Within the approximately 280-mm-long trapping region, we impose roughly 2.25 spatial periods of an electric potential that varies between approximately ± 100 V (see Fig. 1c). These potentials come in fifty time-identical repetitions (see Fig. 2a). Individual repetitions last 1.698 s and are spaced at intervals of 2.298 s. Each repetition begins and ends with the potentials zeroed. Within each repetition, the potential cycles between that shown in Fig. 1c and its negation; the average time interval between transitions is 1 ms. Thus, there are 1,698 transitions in each repetition and $N = 84,900$ transitions in total. Typical time histories of the potentials are shown in Fig. 2b–d. The transitions begin in synchronization for all electrodes; the differences between the attacks (the rise times of each pulse) and magnitudes of the three potentials shown in Fig. 2b–d are due to deliberate differences in the amplifiers and filters driving the electrodes. These differences are precisely modelled in our simulations, and are not otherwise germane to the results presented here.

Data collection. All data were collected in two adjacent half-shifts, separated by an off-shift period, at CERN's AD. The data collection followed the protocol outlined above of strict alternation of stochastic and null trials. Data collection ended when we reached the predetermined number (ten) of trials of each type. Trials that had an immediately obvious failure (no antiprotons delivered from the AD, premature failure of our magnet system and so on) were aborted and repeated; otherwise, all the data that were collected are presented here.

Equation of motion. The antiatom equation of motion in the trap is

$$M\ddot{\mathbf{x}} = \nabla(\boldsymbol{\mu} \cdot \mathbf{B}) + Qe(\mathbf{E} + \dot{\mathbf{x}} \times \mathbf{B}) \quad (2)$$

Here, $\mathbf{x} = \mathbf{x}(t)$ is the centre-of-mass position of an antiatom at time t , $\mathbf{E} = \mathbf{E}(\mathbf{x}, t)$ and $\mathbf{B} = \mathbf{B}(\mathbf{x}, t)$ are the applied electric and magnetic fields, M is the total mass and $\boldsymbol{\mu}$ is the magnetic moment of antihydrogen. A trapped antiatom must be low- B -field-seeking, so the direction of $\boldsymbol{\mu}$ is assumed to adiabatically track the direction antiparallel to $\mathbf{B}(\mathbf{x}, t)$.

Simulations. The simulations mimic the experimental procedures closely, and are similar to those in refs 1, 2, 5, 14, 15, 25 and 26, in which numerous tests of robustness of the dynamical model and numerical approximations have been performed. Simulated antiatoms are launched from a volume corresponding to the precursor positrons. The antiatoms are drawn from an energy distribution obeying a truncated 3D Maxwellian, which scales as $E_0^{1/2}$ at low energies, where E_0 is the initial energy of the antiatoms. The truncation energy is set to 0.75 K, well above the trap depth of $E_{\text{well}} = 0.54$ K, but well below the positron temperature of tens of kelvin that sets the overall Maxwellian temperature. As discussed below, the antiatoms are then held for 1.2 s. Most of the antiatoms with energies above E_{well} will be lost during this relaxation/randomization time; however, a few, on quasi-bound orbits, will remain^{2,29}.

In the experiment, we mix positrons and antiprotons for 1 s. Antihydrogen production peaks at the beginning of the mixing phase, but we do not know when within the mixing phase the antihydrogen that is trapped is created. After mixing, the antiatoms are allowed to randomize for 0.4 s. During this relaxation/randomization phase, electric fields are applied that clear the positrons and antiprotons remaining in the trap. We do not model these fields because they are generally weak compared to the stochastic potentials, there are few transitions and, experimentally, both the stochastic and null trials experience these fields. Altogether, these phases take a total of 1.4 s; in the simulations, to account for the uncertainty in the synthesis time, we allot 1.2 s for this composite relaxation/randomization time. We note that 1.2 s is short compared to the 114.9-s stochastic/hold phase, and that relaxation continues during this phase. Thus, the 0.2-s uncertainty has a negligible effect on our results.

We launch 1,000 antiatoms for each case we study. About 67% of these antiatoms survive the initial relaxation phase. Because 61% have initial energy below E_{well} , the remaining 6% are on quasi-bound orbits.

As discussed in the main text, we then propagate the antiatoms remaining after the relaxation phase for 114.9 s without the stochastic electric fields (null trials), or with the stochastic fields and with an assumed Q (stochastic trials). At the end of the 114.9 s, we count the number of antiatoms that have survived the simulation to determine the survival probability; there is no need to model the magnetic field shut-off used to count the antiatoms in the experiment.

We use a symplectic leap-frog propagator to numerically integrate equation (2). The electric potentials in the simulation are determined from a precise model of the trap geometry, using the COMSOL finite element code³⁰ to solve the Laplace equation, and the measured time-history of the potentials on each electrode (see Fig. 2). The magnetic field is calculated from high-accuracy analytic expansions derived from numerical Biot–Savart modelling²⁵. We do not see any interesting effects from coarsening the field solutions (see also ref. 5).

Simulation convergence studies show that the step size need not be shorter than 4 μs ; we used 0.8 μs for the simulations reported here. They also indicate that the electrode drive amplifier response time need only be known to the 4- μs level, which is well within its measured accuracy.

Bayesian analysis. In the Supplementary Information, we present a complete Bayesian analysis that explicitly accounts for uncertainty in the relationship between s and $|Q|$ due to the finite number of samples in our Monte Carlo simulations, and for various systematic effects. Here, we give an abbreviated analysis. We search for the value of the survival probability, s^* , that yields a $(1 - \alpha)$ posterior credible interval given our observations

$$1 - \alpha = \int_{s^*}^1 ds' p(s' | N_s, N_n) \quad (3)$$

Here, $p(s | N_s, N_n)$ is the probability density for the survival probability s , given a total of N_s observed atoms in the stochastic trials and N_n observed atoms in the null trials, and $1 - \alpha = 0.683$ for the equivalent of a 1σ credible interval. Bayesian formalism gives

$$p(s | N_s, N_n) = \frac{\int_0^\infty dr' p(N_s, N_n | s, r') \pi(s) \pi(r')}{\int_0^1 ds' \int_0^\infty dr' p(N_s, N_n | s', r') \pi(s') \pi(r')} \quad (4)$$

where $\pi(s)$ and $\pi(r)$ are the priors on s and r , respectively. The likelihood $p(N_s, N_n | s, r)$ is given by

$$p(N_s, N_n | s, r) = \sum_{n \geq N_s} \left[\binom{n}{N_s} s^{N_s} (1-s)^{n-N_s} \right] \left[\frac{(Kr)^n}{n!} e^{-Kr} \right] \left[\frac{(Kr)^{N_n}}{N_n!} e^{-Kr} \right] \quad (5)$$

$$= \frac{(Krs)^{N_s}}{N_s!} e^{-Krs} \frac{(Kr)^{N_n}}{N_n!} e^{-Kr} \quad (6)$$

where $K = 10$ is the number of stochastic and null trials. The first factor in the summation in equation (5) is a binomial subsampling of the Poisson-distributed number (given by the second factor) of antiatoms n present in the stochastic trials. (The Poisson assumption is further discussed in the Supplementary Information.) This subsampling models loss, with probability $1 - s$, due to the stochastic fields. The third factor is the Poisson-distributed number of antiatoms present in the null trials. The first two terms can be combined into a single Poisson process, yielding equation (6).

The bound on $|Q|$ compatible with the current experiment is substantially lower than the previously reported bound⁵, and we do not possess accurate prior information about the actual observation rate r in the particular trials used here. Hence, we choose to use uninformative priors. Various notions of uninformative priors have been suggested³¹. For both binomial and Poisson sampling distributions, the widely used Bernardo–Berger reference priors³², Jeffreys invariant priors³³ and Rodriguez entropic priors³⁴ all coincide, and are given by

$$\pi(s) \propto \frac{1}{\sqrt{s(1-s)}}; \quad \pi(r) \propto \frac{1}{\sqrt{r}} \quad (7)$$

When evaluated, equations (3)–(7) suggest that $s > 0.79$ and that the corresponding bound on $|Q|$, found by inverting the curve in Fig. 3, is $|Q| < 0.59$ p.p.b. without considering systematic uncertainties.

Bounding the electric dipole moment. If we take $Q = 0$, but change the equation of motion, equation (2), to

$$M\ddot{\mathbf{x}} = \nabla(\boldsymbol{\mu} \cdot \mathbf{B}) + \alpha \nabla |\mathbf{E}|^2 \quad (8)$$

to include the effects of electric polarizability, then we can reinterpret our experiment as setting a bound on the polarizability α of the antihydrogen atoms (in the absence of appreciable permanent monopole or dipole moments). This sets a limit of $\alpha < 3 \times 10^6 \alpha_0$, at 68.3% confidence, not including systematic effects, where $\alpha_0 = 4\pi\epsilon_0 \times 0.667 \times 10^{-24} \text{ cm}^3$ is the known polarizability of hydrogen⁸. Note that effects due to polarizability will ultimately limit the precision of the stochastic methodology.

If instead we take $Q = 0$ and $\alpha = 0$, and change the equation of motion to

$$M\ddot{\mathbf{x}} = \nabla(\boldsymbol{\mu} \cdot \mathbf{B}) + \nabla(\mathbf{p} \cdot \mathbf{E}) \quad (9)$$

then we can interpret our experiment as setting a bound on a permanent electric dipole moment \mathbf{p} of antihydrogen. This gives a bound of 10^{-9} e cm at 68.3% confidence not including systematic effects. If this moment were to come from, for instance, a positron dipole moment, then quantum corrections³⁵ would loosen

the bound to about $10^{-5}e\text{cm}$. This bound is many orders of magnitude greater than the current bound on the electron dipole moment of $9.7 \times 10^{-29}e\text{cm}$ (ref. 36). **Identification of antihydrogen.** Antihydrogen detection, and cosmic ray rejection, is based on previous work within the ALPHA experiment (see ref. 37). By taking the energy deposited in each strip of the SVD, clusters of activated strips can be summed into 'raw hits'; raw hits from either side of a silicon module can be paired to make 'hits'; hits can be compared in 3D space to create 'tracks'; and tracks can be combined to point to an annihilation vertex.

The detector only detects the pions that are produced by the antiproton annihilations when antihydrogen atoms annihilate on the trap wall or on the very-low-density background gas; the detector is insensitive to the γ -rays produced by the simultaneous positron annihilations. We have developed many techniques for first eliminating antiprotons from our trap before the detection phase, and then discriminating between antihydrogen annihilations and the unlikely antiproton annihilations that might occur if some unbound antiprotons were still present^{1,17,25}. We are confident that the events reported here are due to antihydrogen annihilations only.

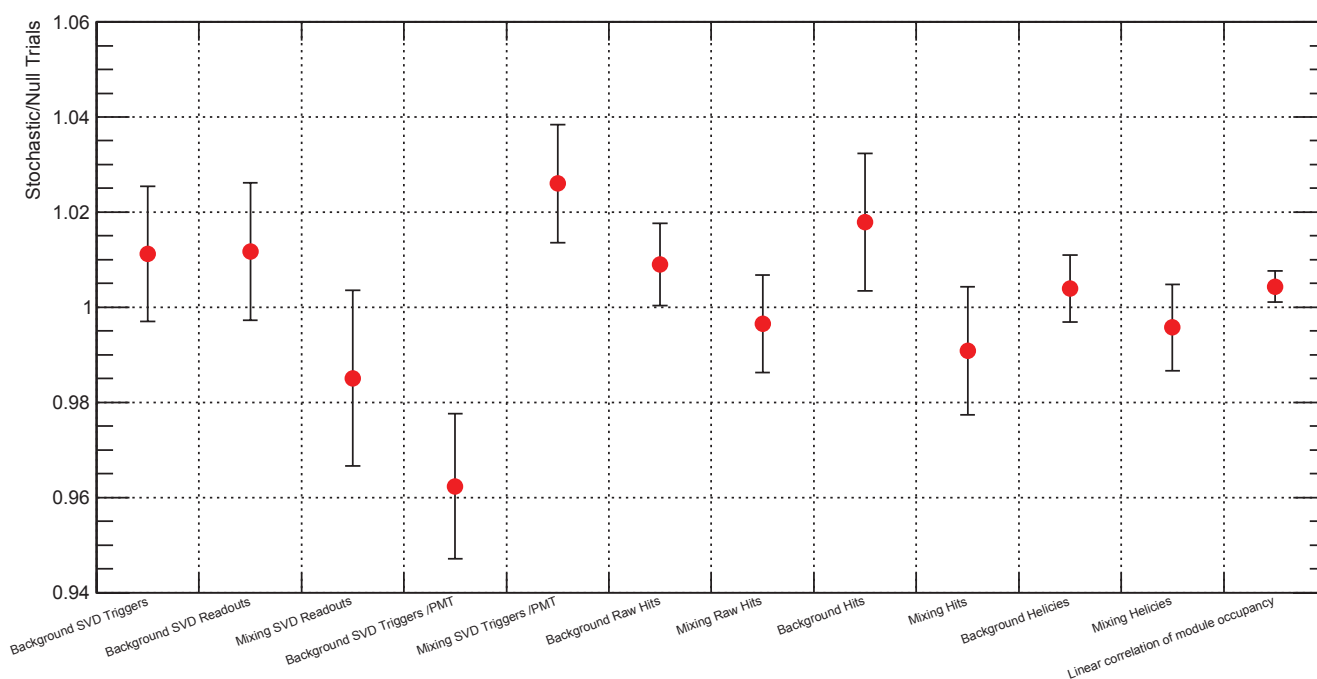
To determine the best significance of antihydrogen detection against background, we fed multiple selection variables that measure the event shape and orientation into a MVA^{22–24}. Complete reconstruction of the annihilation vertex is only part of the input to this MVA, because this method combines low- and high-level reconstruction data in a machine-learning algorithm. We used a large annihilation data sample and a large, independent, background data sample to train the machine-learning algorithm. The annihilation samples have fewer than 1% cosmic events and the background sample consists entirely of background events because the data are collected in the absence of antiprotons in the experiment. We then tested the performance of the MVA technique on a dedicated, separate test sample of data. Once trained, we used the MVA to analyse events that were detected during the experimental trials, and classified each event as either an antihydrogen annihilation event or as a background cosmic event. Note that we can tune the MVA differently for significance optimization during the stochastic/hold phase (2) and the release phase (3); for instance, we can decrease the false-negative rate where possible (increase the detection efficiency), at the cost of increasing the false-positive rate.

MVA of the release phase (3). With the MVA optimized for the 30-ms observation window of the release phase (3), $(82.3 \pm 0.2)\%$ of the annihilation events are accepted with a background rate of $242\text{ mHz} \pm 2\%$, which gives $(0.073 \pm 0.002) \ll 1$ expected cosmic rays in the observation window of both the null and stochastic trials. The results of these trials are reported in Table 1, and are the primary data used here.

MVA of the stochastic/hold phase (2). For the 10 null and 10 stochastic trials, events detected during the stochastic/hold phase (2) (before release) were fed into a differently trained MVA classifier that has a detector acceptance of $(46.1 \pm 0.3)\%$ and a cosmic background rate of 6 mHz with 11% uncertainty (see Extended Data Table 1). Stochastic/hold phase events might have arisen from the escape of quasi-bound antiatoms, from annihilations on residual gas, or from ejection of the antiatoms by the stochastic acceleration. We performed a detailed statistical analysis to determine whether anything other than background needs to be introduced to account for the observed counts in Extended Data Table 1. That analysis (see Supplementary Information) shows that there is no reason to reject the assumption that these phase (2) counts are due to background, and thus do not provide reliable additional information regarding Q.

Detector performance. Because consistency between the detector performance during the null and stochastic trials is critical for this analysis, we rigorously scrutinized and cross-checked the performance of the SVD with scintillation detectors placed on the outside of the apparatus. We examined several quality-of-data (QOD) criteria using both cosmic data from dedicated cosmic observation windows, and annihilation data from the mixing phase of experimental trials. We found no anomalous differences between the stochastic and null trials. The results of these tests²⁷ are summarized in Extended Data Fig. 1.

29. Bowman, J. D. & Penttila, S. I. On the measurement of the neutron lifetime using ultracold neutrons in a vacuum quadrupole trap. *J. Res. Natl. Inst. Stand. Technol.* **110**, 361–366 (2005).
30. COMSOL, <http://www.comsol.com/> (COMSOL, Inc., 2015).
31. Kass, R. E. & Wasserman, L. The selection of prior distributions by formal rules. *J. Am. Stat. Assoc.* **91**, 1343–1370 (1996).
32. Berger, J. O. *Statistical Decision Theory and Bayesian Analysis* 2nd edn (Springer, 1985).
33. Jeffreys, H. *Theory of Probability* 3rd edn (Clarendon Press, 1961).
34. Rodriguez, C. C. in *Maximum Entropy and Bayesian Methods* (ed. Fougère, P. F.) Vol. 39 of *Fundamental Theories of Physics* 31–39 (Springer, 1990).
35. Sandars, P. G. H. The electric-dipole moments of an atom II. The contribution from an electric-dipole moment on the electron with particular reference to the hydrogen atom. *J. Phys. B* **1**, 511–520 (1968).
36. The ACME Collaboration. Order of magnitude smaller limit on the electric dipole moment of the electron. *Science* **343**, 269–272 (2014).
37. Andresen, G. B. *et al.* Antihydrogen annihilation reconstruction with the ALPHA silicon detector. *Nucl. Instrum. Methods Phys. Res. A* **684**, 73–81 (2012).



Extended Data Figure 1 | Quality-of-data (QOD) tests. QOD tests comparing the ratios of various quantities during stochastic and null trials. No difference corresponds to a ratio of 1, and the error bars are 1σ . Both mixing and background data were analysed in several tests. Background SVD triggers refers to the number of times the SVD meets our trigger condition. This number is counted even when the detector is vetoed because it is being read out. Background/mixing SVD readouts refers to the number of events read out by the detector. During cosmic (background) data collection, this value is not saturated; however, during mixing it is. Background/mixing SVD triggers/PMT is the ratio of the number of readout events that trigger the SVD, and the number of

counts in adjacent scintillating detectors³. The scintillators have a lower detection efficiency and faster response than the SVD, so this ratio tests for saturation of the SVD rate. Background/mixing raw hits is a count of strip clusters with energy deposited 3σ above noise. This is the lowest level of reconstruction analysis. Background/mixing hits is a count of voxel hit clusters, that is, the number of coincident raw hits on both the n- and p-side of the silicon detector modules. Background/mixing helices is the number of three-hit combination tracks that point to the centre of the SVD. Linear correlation of module occupancy is the gradient of a line of best fit between the relative occupancy counts in the two experiments. Here occupancy is defined as the number of hits in a module.

Extended Data Table 1 | Results of MVA during the stochastic/hold phase (2)

Trial Type	Events
Stochastic Trials	6
Null Trials	11
Predicted Background	6.9

See Table 1 for the observed events during the release phase (3).

A lithium–oxygen battery based on lithium superoxide

Jun Lu^{1*}, Yun Jung Lee^{2*}, Xiangyi Luo^{3,4*}, Kah Chun Lau^{3*}, Mohammad Asadi^{5*}, Hsien-Hau Wang³, Scott Brombosz³, Jianguo Wen⁶, Dengyun Zhai¹, Zonghai Chen¹, Dean J. Miller⁶, Yo Sub Jeong², Jin-Bum Park², Zhigang Zak Fang⁴, Bijandra Kumar⁷, Amin Salehi-Khojin⁵, Yang-Kook Sun², Larry A. Curtiss³ & Khalil Amine¹

Batteries based on sodium superoxide and on potassium superoxide have recently been reported^{1–3}. However, there have been no reports of a battery based on lithium superoxide (LiO₂), despite much research^{4–8} into the lithium–oxygen (Li–O₂) battery because of its potential high energy density. Several studies^{9–16} of Li–O₂ batteries have found evidence of LiO₂ being formed as one component of the discharge product along with lithium peroxide (Li₂O₂). In addition, theoretical calculations have indicated that some forms of LiO₂ may have a long lifetime¹⁷. These studies also suggest that it might be possible to form LiO₂ alone for use in a battery. However, solid LiO₂ has been difficult to synthesize in pure form¹⁸ because it is thermodynamically unstable with respect to disproportionation, giving Li₂O₂ (refs 19, 20). Here we show that crystalline LiO₂ can be stabilized in a Li–O₂ battery by using a suitable graphene-based cathode. Various characterization techniques reveal no evidence for the presence of Li₂O₂. A novel templating growth mechanism involving the use of iridium nanoparticles on the cathode surface may be responsible for the growth of crystalline LiO₂. Our results demonstrate that the LiO₂ formed in the Li–O₂ battery is stable enough for the battery to be repeatedly charged and discharged with a very low charge potential (about 3.2 volts). We anticipate that this discovery will lead to methods of synthesizing and stabilizing LiO₂, which could open the way to high-energy-density batteries based on LiO₂ as well as to other possible uses of this compound, such as oxygen storage.

The crystalline LiO₂ reported here was made electrochemically using a cathode based on reduced graphene oxide (rGO) with added iridium (Ir) nanoparticles. Initially graphene oxide (GO) was prepared by a modified Hummer's method^{21,22}. The Ir–rGO composite was then made by a hydrothermal reduction method and characterized (Supplementary Fig. 2). Scanning electron microscopy (SEM) images of the pristine rGO and Ir–rGO composite (Fig. 1a and b, respectively) reveal porous three-dimensional (3D) networks of rGO composed of wrinkled 2D rGO sheets. Figure 1c and d shows transmission electron microscopy (TEM) images of the Ir nanoparticles on rGO, indicating that the well-dispersed Ir nanoparticles decorated on rGO are very small (<2 nm), with evidence for the presence of some small Ir clusters (circled in Fig. 1d). A backscattering image (Supplementary Fig. 1) shows some scattered larger Ir particles of about 500 nm in size, which may be due to agglomeration of the smaller nanoparticles, and fast Fourier transform analysis of high-resolution (HR)-TEM images (Supplementary Fig. 1) show that the nanoparticles are Ir. An X-ray photoemission spectroscopy (XPS) analysis (Supplementary Fig. 1) indicates the Ir surface is only partially oxidized.

The performance of the rGO and Ir–rGO cathodes was examined using a Swagelok-type cell composed of a lithium metal anode,

electrolyte (1 M LiCF₃SO₃ in tetraethylene glycol dimethyl ether (TEGDME)) impregnated into a glass fibre separator, and a porous cathode. A current density of 100 mA g^{−1} was used for both discharge and charge, and the cell was run with a capacity limit of 1,000 mA h g^{−1} to avoid side reactions. The specific capacity (mA h g^{−1}) and the current density (mA g^{−1}) are based on the active materials of the O₂ electrodes. Figure 2a and b shows voltage profiles for the Ir–rGO and rGO cathode architectures, respectively. The Ir–rGO discharge product shows a very low charge potential of ~3.2 V that rises to 3.5 V over 40 cycles leading to more than 85% efficiency in this system (Fig. 2a). The voltage profile of the rGO cathode shows a much larger charge potential of ~4.2 V with a lower efficiency of ~67% (Fig. 2b).

The discharge product resulting from the Ir–rGO cathode was examined using SEM, differential electrochemical mass spectroscopy (DEMS), high-energy X-ray diffraction (HE-XRD), TEM and Raman spectroscopy with the results shown in Figs 2 and 3. The SEM image in Fig. 2c shows the Ir–rGO cathode after discharge (~2.75 V) from the first cycle (1,000 mA h g^{−1} capacity). This image indicates that the discharge product resulting from the Ir–rGO-based cathode consists mainly of nanoparticles with needle- or rod-like morphology, although the presence of other shapes such as cubic cannot be ruled out. This needle- or rod-like morphology is also observed in the TEM image of a part of the discharge product, which appears to be on the surface of the Ir–rGO nanostructures (Fig. 2c inset). An SEM image after charging shows that the nanoparticles have disappeared (Supplementary Fig. 3). The discharge product from the rGO-based cathode has a range of morphologies, including toroids and nanoparticles (Supplementary Fig. 3). The Ir–rGO discharge product produced by a current density of 100 mA g^{−1} was characterized by DEMS during the first charging cycle by on-line monitoring of the number of evolved O₂ molecules. The experiment was performed using high current densities (1,000 mA g^{−1} and 640 mA g^{−1}) for charging to enable measurement of the evolved O₂. The DEMS results at the higher current density are shown in Fig. 2d. Analysis of the data in Fig. 2d gives an average O₂ formation rate of 1.3×10^{-9} mol s^{−1}, resulting in an e[−]/O₂ ratio of 1.00. A similar (1.00) e[−]/O₂ ratio was also obtained for the experiment with a current density of 640 mA g^{−1} (Supplementary Fig. 18). Additionally, negligible amounts of CO₂ and H₂ gases were generated during the DEMS experiments (Fig. 2d and Supplementary Fig. 18). A DEMS experiment was also carried out during discharge, and gave an e[−]/O₂ ratio of 1.02 (Supplementary Fig. 18, Supplementary Table 3). These results are consistent with LiO₂ as the main discharge product, and provide evidence for the absence of other products (for example, Li₂O₂, LiOH and Li₂CO₃). The DEMS results for LiO₂ are similar to DEMS measurements on a NaO₂ battery that gave an e[−]/O₂ ratio of 1.00 for discharge and 1.02 for charge².

¹Chemical Sciences and Engineering Division, Argonne National Laboratory, Argonne, Illinois 60439, USA. ²Department of Energy Engineering, Hanyang University, Seoul 133-791, South Korea.

³Materials Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA. ⁴Department of Metallurgical Engineering, University of Utah, Salt Lake City, Utah 84112, USA.

⁵Department of Mechanical and Industrial Engineering, University of Illinois at Chicago, Chicago, Illinois 60607, USA. ⁶Center for Nanoscale Materials, Argonne National Laboratory, Argonne, Illinois 60439, USA. ⁷Conn Center for Renewable Energy Research, University of Louisville, Louisville, Kentucky 40292, USA.

*These authors contributed equally to this work.

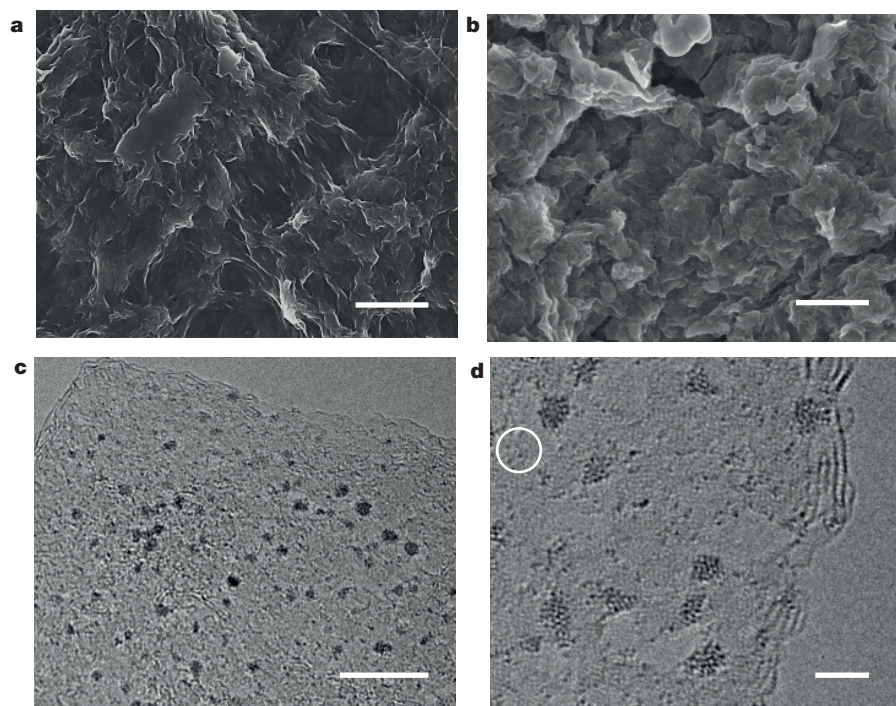


Figure 1 | Morphology of Ir-rGO. **a**, SEM image of pristine rGO powder. **b**, SEM image of Ir-rGO composite. **c**, **d**, TEM images of Ir-rGO composite, showing Ir nanoparticles less than 2 nm in size. The circle in

d shows some small Ir atomic clusters. Scale bars: **a**, **b**, 1 μm ; **c**, 10 nm; **d**, 2 nm. There are also some scattered large Ir agglomerates on the rGO (see Supplementary Fig. 1).

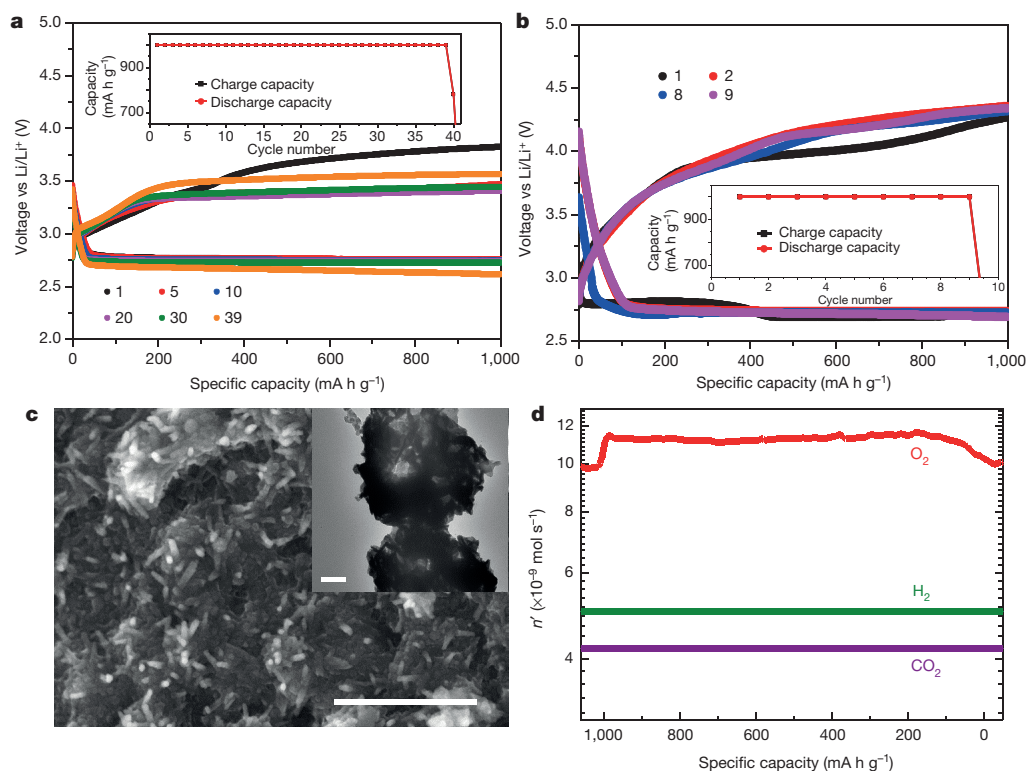


Figure 2 | Electrochemical tests and discharge products. **a**, Voltage profiles of the Ir-rGO cathode (see also Supplementary Fig. 6). Cycle number of voltage plot is given by the colour of the plotting symbol (see key). Inset shows capacity as a function of cycle number. **b**, Voltage profiles of the rGO cathode. Cycle number of voltage plot is given by the colour of the plotting symbol (see key). Inset shows capacity as a function of cycle number. **c**, Main panel, SEM image of discharge product on Ir-rGO (scale bar, 1 μm); inset, TEM image of discharge product on

Ir-rGO (scale bar, 200 nm). Both main panel and inset are from the first discharge with a current density of 100 mA g^{-1} , and capacity control of 1,000 mAh g^{-1} . **d**, DEMS profile showing O_2 , H_2 and CO_2 gases released (n' is the number of moles per second) from the cell during the charging process (at 1,000 mA g^{-1} current density) after the first discharge (at 100 mA g^{-1} current density) to 1,000 mAh g^{-1} capacity (see Supplementary Information for details).

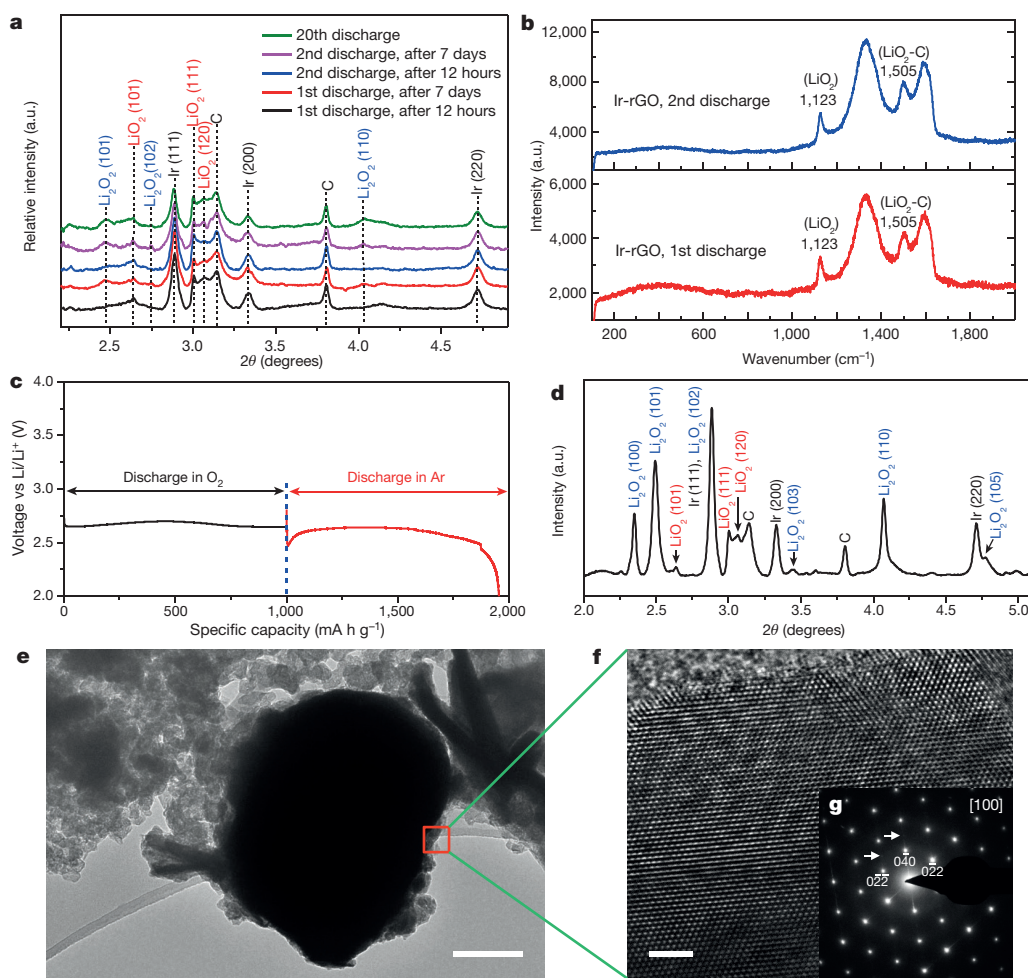


Figure 3 | Characterization of discharge products. **a**, HE-XRD patterns of discharge product on Ir-rGO as a function of ageing time. **b**, Raman spectra of discharge product on Ir-rGO for first and second discharges. **c**, Voltage plots for Ir-rGO discharged first in O_2 to a capacity of $1,000 \text{ mA h g}^{-1}$ and then discharged in Ar during which time it attained a capacity 956 mA h g^{-1} ; in the latter case, the electrolyte was purged with Ar before discharge (see Supplementary Fig. 16 for comparison with other

cathode materials). **d**, HE-XRD pattern of cathode resulting from both discharges in **c**. **e–g**, TEM image of an Ir agglomerate after first discharge (**e**; scale bar, 200 nm); **f**, HR-TEM image of boxed area in **e** (scale bar, 2 nm); and **g**, the corresponding electron diffraction pattern along the $[100]$ zone axis giving evidence for the formation of an Ir_3Li intermetallic. The indices are diffraction vectors. Weak superstructure is observed as indicated by arrows.

The HE-XRD pattern in Fig. 3a for the discharge product on the Ir-rGO cathode ($1,000 \text{ mA h g}^{-1}$ capacity) during the first cycle shows peaks corresponding to crystalline LiO_2 ((101), (111), (120)), and no evidence for peaks corresponding to Li_2O_2 . The identification of the LiO_2 peaks is based on a theoretical XRD pattern derived from the DFT (density functional theory)-predicted crystalline LiO_2 structure (Supplementary Fig. 4) from refs 19 and 23, as no experimental XRD pattern has been reported. The LiO_2 structure is orthorhombic (Supplementary Fig. 4). For comparison, NaO_2 is cubic at room temperature and orthorhombic at $<196 \text{ K}$, whereas KO_2 is tetragonal at room temperature. Some amorphous LiO_2 cannot be ruled out on the basis of the XRD results. The standard XRD pattern of Li_2O_2 was used to determine the absence of Li_2O_2 . The Raman spectra of the discharge product of the Ir-rGO cathode in Fig. 3b show the presence of a peak at $1,123 \text{ cm}^{-1}$, consistent with the range of values that have been observed for superoxide stretching frequencies (Supplementary Table 1). It is also consistent with the Raman peak at $1,156 \text{ cm}^{-1}$ observed¹ for NaO_2 . There is also a peak at $1,505 \text{ cm}^{-1}$ that has recently been attributed to the strong interaction between LiO_2 and a graphitic carbon surface⁹. In contrast, the HE-XRD pattern (Supplementary Fig. 3) for the discharge product on the rGO cathode without Ir added ($1,000 \text{ mA h g}^{-1}$ capacity) during the first cycle shows peaks corresponding to both crystalline LiO_2 ((101), (111), (120)) and Li_2O_2 ((101), (102), (103), (110)). Evidence for both LiO_2 and Li_2O_2

components in the discharge product has also been reported in other studies^{9–15}, although none are based on XRD characterization, which is made possible by the use of high-energy X-rays at the Advanced Photon Source of Argonne National Laboratory. When the Ir-rGO cell is run to deep discharge of 2.2 V and $\sim 9,500 \text{ mA h g}^{-1}$ capacity, the HE-XRD data shows evidence for the presence of LiO_2 , Li_2O_2 and $LiOH$ with a toroidal morphology (Supplementary Fig. 5). We also note that there have been some previous studies^{24,25} on rGO and rGO with Au nanoparticles that showed formation of Li_2O_2 , but no report of LiO_2 in a Li– O_2 cell.

Further evidence that the discharge product is LiO_2 on the Ir-rGO cathode was obtained by an experiment in which Li was electrochemically added to the discharge product without the presence of O_2 (that is, O_2 was replaced by Ar). The voltage profile is shown in Fig. 3c for this discharge process, along with that of the initial discharge process (to $1,000 \text{ mA h g}^{-1}$). The HE-XRD of the resulting product with no O_2 in the cell is shown in Fig. 3d and reveals strong peaks from Li_2O_2 , thus indicating a conversion of LiO_2 to Li_2O_2 ($Li^+ + e^- + LiO_2 \rightarrow Li_2O_2$) with $\sim 96\%$ of the theoretical capacity for this reaction attained. This is evidence for a reaction involving one electron per O_2 in the first cycle for $1,000 \text{ mA h g}^{-1}$ capacity, and for no crystalline or amorphous Li_2O_2 forming on the initial capacity-limited discharge. In contrast, no significant capacity for discharge in Ar is observed for the rGO cathode when a similar procedure is performed. We have also carried

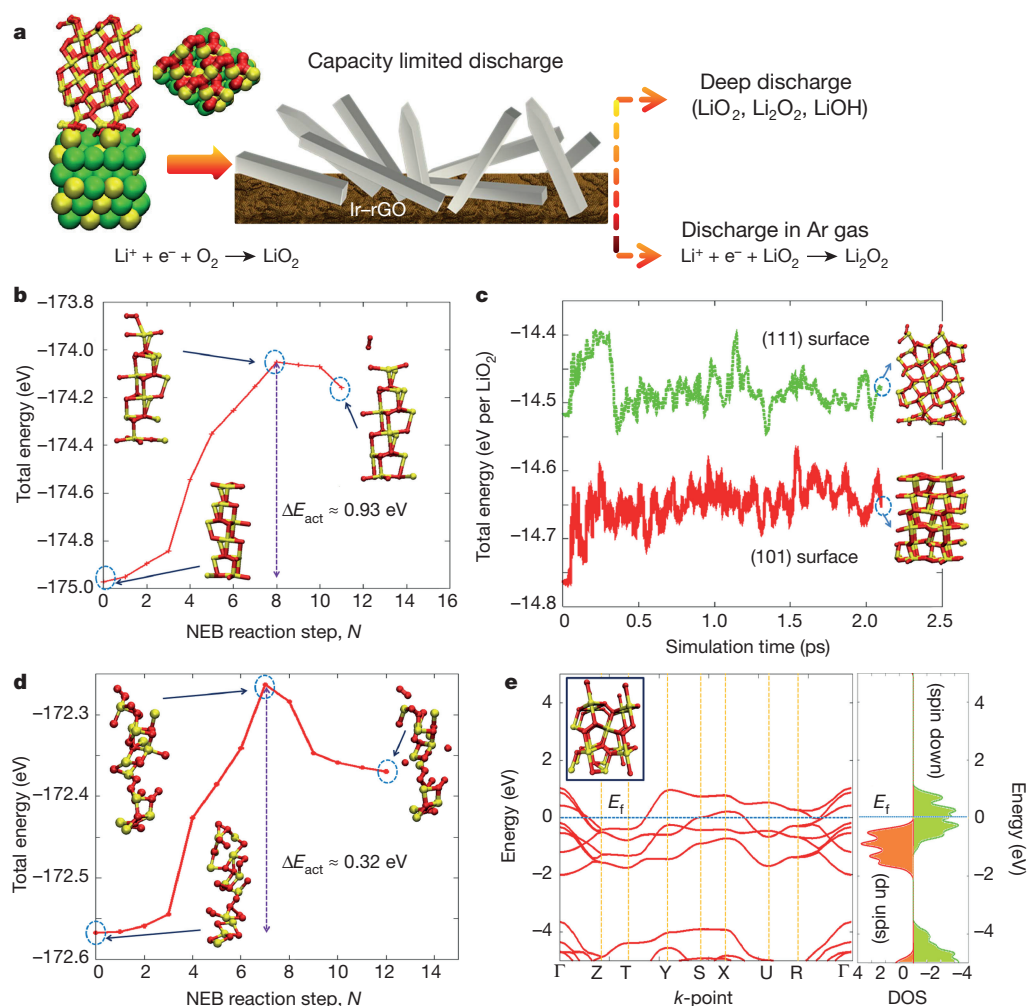


Figure 4 | Density functional calculations. **a**, Schematic showing lattice match between LiO_2 and Ir_3Li (see also Supplementary Fig. 13) that may be responsible for the LiO_2 discharge product found on the Ir-rGO cathode. The two structures at left are the side- and top-views representing epitaxial growth of crystalline LiO_2 in (111) orientation on a (121) facet of Ir_3Li (Li is yellow, O is red and Ir is green). The rod-like structures shown in the central figure are schematic representations of the crystalline LiO_2 morphology observed in the experiment. Two subsequent electrochemical reactions that the LiO_2 can undergo are shown on the right: that is, either further lithiation in the presence of Ar or further oxygen reduction in

the presence of oxygen. **b**, DFT calculations of the barrier for desorption (ΔE_{act}) of an O_2 molecule from the (101) LiO_2 surface in vacuum from a Nudged Elastic Band (NEB) calculation as a function of number of reaction steps, N . **c**, AIMD simulations of (111) LiO_2 and (101) LiO_2 surfaces in vacuum at room temperature. **d**, DFT calculations of the barrier for desorption of an O_2 molecule from an amorphous LiO_2 surface in vacuum. **e**, DFT electronic band structure (left) and density of states (DOS) plot (right) of ferromagnetic bulk crystalline LiO_2 close to the Fermi level (E_f) based on a spin polarized calculation with electronic spin-up and spin-down states shown. (See also Supplementary Fig. 17.)

out electron paramagnetic resonance (EPR) measurements, and find a signal that is consistent with the presence of LiO_2 (Supplementary Fig. 7, Supplementary Table 2) in the initial discharge product.

The stability of the LiO_2 was investigated by carrying out HE-XRD measurements on Ir-rGO cathodes aged for different times in the presence of the electrolyte (Fig. 3a). After 12 hours at the end of both the first and second discharges the XRD patterns show only evidence for crystalline LiO_2 . When the discharge product is allowed to remain for seven days under the same conditions, both samples still show the signature of LiO_2 along with the presence of crystalline Li_2O_2 . Thus, the HE-XRD measurements indicate that the crystalline LiO_2 formed with the Ir-rGO cathode is surprisingly stable for a relatively long period of time. Raman spectra (Supplementary Fig. 8) measured as a function of time exhibit a decreasing intensity of the peak at $1,123\text{ cm}^{-1}$ as a function of time, which is consistent with the HE-XRD results. In addition, LiO_2 is still the dominant discharge product on the twentieth discharge cycle, indicating that LiO_2 is stable enough that it can be repeatedly charged and discharged for about 40 cycles with a very low charge potential (~ 3.2 V); this may open the way to a lithium-superoxide-based battery.

An explanation of the formation mechanism and stability of the LiO_2 found in this study requires an understanding of the growth and nucleation process, which is quite complex and beyond the scope of the present study. However, our results are consistent with mechanisms proposed in other recent studies. Some insight can be obtained from a postulated mechanism for the nucleation and growth of discharge products from various size-specific Ag clusters decorating carbon cathodes in a Li- O_2 cell⁷. In that case, the results are explained by a through-solution growth mechanism with sites for oxygen reduction reactions (ORRs) that are separate from the nucleation sites. Other researchers have also reported evidence for through-solution mechanisms^{26–29}. In addition, the results of the Ag cluster study⁷ suggest that availability of good ORR sites promotes a through-solution mechanism involving nucleation and growth of LiO_2 followed by disproportionation to Li_2O_2 . Since both Ir and rGO are good ORR materials^{30,31}, the Ir-rGO-based cathodes of the present study should result in a similar mechanism, that is, initial formation of LiO_2 .

The HE-XRD finding that rGO results in both LiO_2 and Li_2O_2 in the discharge product can be accounted for by slow disproportionation. This explanation is supported by a recent theoretical study¹⁷ that has

shown the possibility of fast and slow disproportionation processes for LiO_2 depending on the LiO_2 cluster size. In addition, experimental evidence for slow disproportionation has been found in studies of Li-O_2 cell discharge products based on an activated carbon cathode¹⁰. In contrast, the present HE-XRD results showing that only LiO_2 is present in the case of the Ir-rGO-based cathode material suggest that disproportionation is suppressed.

The formation of only LiO_2 in the case of the present Ir-rGO cathode may be due to some aspect of that cathode that favours nucleation and growth of largely the crystalline LiO_2 phase, which prevents disproportionation (see theoretical calculations below). In the case of Ir-rGO, we noticed the formation of an Ir_3Li intermetallic compound on the large Ir agglomerates seen in the backscattering image (Supplementary Fig. 1), as shown Fig. 3f and g). We also noticed that some nanoparticles (needle or rod-like) are formed on the surface of these agglomerates during the first discharge (Fig. 3e). We note that the Ir_3Li intermetallic compound has an orthorhombic lattice, a similar crystallographic lattice³² to that of LiO_2 . It is possible that it may act as a template for growth of the crystalline LiO_2 , as has been found in template-controlled nucleation and growth of other crystalline materials³³. We carried out DFT calculations on the interface between LiO_2 and Ir_3Li , and found that some crystalline faces had good lattice matches (Supplementary Fig. 13), as would be required for epitaxial growth of crystalline LiO_2 . There may be other intermetallic compounds that exhibit similar behaviour and will be the subject of further study. The schematic in Fig. 4 summarizes the novel templating process that may be responsible for the LiO_2 discharge product found for the Ir-rGO cathode material and the subsequent electrochemical reactions that it can undergo, that is, either further lithiation or further oxygen reduction. In contrast, the rGO cathode, which does not include Ir nanoparticles, probably has a different nucleation and growth mechanism resulting in a discharge product composed of both LiO_2 and Li_2O_2 .

The kinetic stability of crystalline and amorphous LiO_2 was investigated using *ab initio* molecular dynamics (AIMD) and DFT calculations with the results shown in Fig. 4. The disproportionation rate will depend on several factors. One factor is the rate at which the O_2 leaves the surface. The DFT results in Fig. 4b indicate that the initial step of O_2 leaving the crystalline surface into vacuum has a barrier of ~ 0.9 eV based on a low-energy LiO_2 surface. Figure 4c shows that crystalline LiO_2 surfaces (that is, (101) and (111)) are thermally stable in vacuum at room temperature. For an amorphous surface, the barrier (~ 0.3 eV) is less than for the crystalline surface (Fig. 4d). From AIMD simulations, the presence of some solvent molecules adsorbed on the amorphous LiO_2 surface reduces O_2 desorption (Supplementary Fig. 14). This suggests that solvent on the LiO_2 surface could further suppress disproportionation of the crystalline phase. The electrolyte effect on disproportionation was investigated by allowing a sample from a $1,000 \text{ mA h g}^{-1}$ discharge to age for 24 h in vacuum. Characterization of the sample by Raman spectroscopy, discharge in Ar, and charge potential shows a significant decrease of LiO_2 signature after ageing in vacuum, indicating that kinetics plays an important role in stabilizing the LiO_2 (Supplementary Fig. 15).

The Ir-rGO cathode also exhibits a low charge potential, which may be due to several factors. As shown in Fig. 4e, crystalline LiO_2 is a half-metal (on the basis of density functional calculations) and, thus, will have good electronic conduction, in contrast to insulating bulk Li_2O_2 . Another factor is that Ir is known to be a good oxygen evolution catalyst³¹ and interacts strongly with LiO_2 to form a good interface for electrical contact. These properties may explain why the discharge product formed on just rGO has a large charge potential, that is, it lacks the Ir nanoparticles. The Li-O_2 cell based on Ir-rGO cathode material also can cycle 40 or more times (Supplementary Fig. 9) before failure, similar to what has been found for Li_2O_2 -based Li-O_2 cells, indicating that the lithium superoxide is not any more reactive towards the electrolyte than lithium peroxide. In addition, the low charge potential

will lead to less side reactions. The failure of the cell could be due to oxygen crossover to the anode resulting in the anode being converted to LiOH , as evidenced by the corrosion of the anode (Supplementary Fig. 10) and, possibly the poisoning of Ir metal catalyst with cycling. When the cycled Li anode is replaced by a new anode, the cell cycles another 30 times (Supplementary Fig. 10).

The evidence presented here indicates that a Li-O_2 electrochemical cell based on a LiO_2 discharge product is possible with a reasonable cycle life, very high efficiency, and a good capacity. The performance characteristics of the cell based on LiO_2 are comparable to those of previously reported electrochemical cells based on KO_2 (ref. 3) and on NaO_2 (ref. 1), although some aspects—such as the charge and discharge potentials—differ. Problems with electrolyte stability and decomposition, as for the electrolytes used for other Li-O_2 systems, probably still remain, but they do not seem any worse than for those systems. There is little evidence of any side reactions in the Raman data for the first discharge cycle (Fig. 3b), or from Raman and Fourier transform infrared data after charging for up to 30 cycles (Supplementary Figs 11, 12), or from NMR data up to 20 cycles (Supplementary Fig. 12), although there could be decomposition products that are not detected. The Fourier transform infrared and Raman results also confirm that the discharge product is not present after charging.

In summary, we have reported evidence it is possible to have a one-electron discharge process that forms only LiO_2 in a Li-O_2 electrochemical cell. This is different from the previous studies^{9–15} that have provided evidence for both LiO_2 and Li_2O_2 in the discharge product of Li-O_2 batteries with some cathode and electrolyte materials, and from studies¹⁶ that have shown LiO_2 can be present in solution during discharge. The evidence for the existence of the LiO_2 comes from DEMS and HE-XRD data with no evidence for Li_2O_2 being present. The results of TEM and density functional calculations indicate that a novel templating growth mechanism involving the use of Ir nanoparticles may be responsible for the crystalline LiO_2 growth. The LiO_2 formed in this way is stable enough to be repeatedly charged and discharged with a very low charge overpotential.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 October 2014; accepted 13 November 2015.

Published online 11 January 2016.

- Hartmann, P. *et al.* A rechargeable room-temperature sodium superoxide (NaO_2) battery. *Nature Mater.* **12**, 228–232 (2012).
- Hartmann, P. *et al.* A comprehensive study on the cell chemistry of the sodium superoxide (NaO_2) battery. *Phys. Chem. Chem. Phys.* **15**, 11661–11672 (2013).
- Ren, X. & Wu, Y. A low-overpotential potassium–oxygen battery based on potassium superoxide. *J. Am. Chem. Soc.* **135**, 2923–2926 (2013).
- Lu, J. *et al.* A nanostructured cathode architecture for low charge overpotential in lithium–oxygen batteries. *Nature Commun.* **4**, 2383 (2013).
- Bruce, P. G., Freunberger, S. A., Hardwick, L. J. & Tarascon, J.-M. Li-O_2 and Li-S batteries with high energy storage. *Nature Mater.* **11**, 19–29 (2011).
- Lu, J. *et al.* Aprotic and aqueous Li-O_2 batteries. *Chem. Rev.* **114**, 5611–5640 (2014).
- Lu, J. *et al.* Effect of the size-selective silver clusters on lithium peroxide morphology in lithium–oxygen batteries. *Nature Commun.* **5**, 4895 (2014).
- Black, R., Lee, J.-H., Adams, B., Mims, C. A. & Nazar, L. F. The role of catalysts and peroxide oxidation in lithium–oxygen batteries. *Angew. Chem. Int. Edn* **52**, 392–396 (2013).
- Zhai, D. *et al.* Raman evidence for late stage disproportionation in a Li-O_2 battery. *J. Phys. Chem. Lett.* **5**, 2705–2710 (2014).
- Zhai, D. *et al.* Disproportionation in Li-O_2 batteries based on a large surface area carbon cathode. *J. Am. Chem. Soc.* **135**, 15364–15372 (2013).
- Yang, J. *et al.* Evidence for lithium superoxide-like species in the discharge product of a Li-O_2 battery. *Phys. Chem. Chem. Phys.* **15**, 3764–3771 (2013).
- Gittleston, F. S., Ryu, W.-H. & Taylor, A. D. Operando observation of the gold–electrolyte interface in Li-O_2 batteries. *ACS Appl. Mater. Interfaces* **6**, 19017–19025 (2014).
- Gittleston, F. S. *et al.* Raman spectroscopy in lithium–oxygen battery systems. *ChemElectroChem* **2**, 1446–1457 (2015).

14. Olivares-Marín, M. *et al.* Spatial distributions of discharged products of lithium-oxygen batteries revealed by synchrotron X-ray transmission microscopy. *Nano Lett.* **15**, 6932–6938 (2015).
15. Ryu, W.-H., Gittleston, F. S., Schwab, M., Goh, T. & Taylor, A. D. A mesoporous catalytic membrane architecture for lithium-oxygen battery systems. *Nano Lett.* **15**, 434–441 (2015).
16. Schaltin, S., Vanhoute, G., Wu, M., Bardé, F. & Fransaer, J. A QCM study of ORR-OER and an in situ study of a redox mediator in DMSO for Li-O₂ batteries. *Phys. Chem. Chem. Phys.* **17**, 12575–12586 (2015).
17. Das, U., Lau, K. C., Redfern, P. C. & Curtiss, L. A. Structure and stability of lithium superoxide clusters and relevance to Li-O₂ batteries. *J. Phys. Chem. Lett.* **5**, 813–819 (2014).
18. Sangster, J. & Pelton, A. D. The Li-O (lithium-oxygen) system. *J. Phase Equilibria* **13**, 296–299 (1992).
19. Lau, K. C., Curtiss, L. A. & Greeley, J. Density functional investigation of the thermodynamic stability of lithium oxide bulk crystalline structures as a function of oxygen pressure. *J. Phys. Chem. C* **115**, 23625–23633 (2011).
20. Kang, S., Mo, Y., Ong, S. P. & Ceder, G. A facile mechanism for recharging Li₂O₂ in Li-O₂ batteries. *Chem. Mater.* **25**, 3328–3336 (2013).
21. Hummers, W. S. & Offeman, R. E. Preparation of graphitic oxide. *J. Am. Chem. Soc.* **80**, 1339 (1958).
22. Xu, Y., Sheng, K., Li, C. & Shi, G. Self-assembled graphene hydrogel via a one-step hydrothermal process. *ACS Nano* **4**, 4324–4330 (2010).
23. Zhuravlev, Y. N. & Obolonskaya, O. S. Structure, mechanical stability, and chemical bond in alkali metal oxides. *J. Struct. Chem.* **51**, 1005–1013 (2010).
24. Kumar, S., Selvaraj, C., Munichandraiah, N. & Scanlon, L. G. Gold nanoparticles anchored reduced graphene oxide as catalyst for oxygen electrode of rechargeable Li-O₂ cells. *RSC Adv.* **3**, 21706–21714 (2013).
25. Selvaraj, C., Kumar, S., Munichandraiah, N. & Scanlon, L. G. Reduced graphene oxide-polypyrrole composite as a catalyst for oxygen electrode of high rate rechargeable Li-O₂ cells. *J. Electrochem. Soc.* **161**, A554–A560 (2014).
26. Oh, S. H., Black, R., Pomerantseva, E., Lee, J.-H. & Nazar, L. F. Synthesis of a metallic mesoporous pyrochlore as a catalyst for lithium-O₂ batteries. *Nature Chem.* **4**, 1004–1010 (2012).
27. Ren, X., Zhang, S. S., Tran, D. T. & Read, J. Oxygen reduction reaction catalyst on lithium/air battery discharge performance. *J. Mater. Chem.* **21**, 10118–10125 (2011).
28. Mitchell, R. R., Gallant, B. M., Shao-Horn, Y. & Thompson, C. V. Mechanisms of morphological evolution of Li₂O₂ particles during electrochemical growth. *J. Phys. Chem. Lett.* **4**, 1060–1064 (2013).
29. Ottakam Thotiyil, M. M., Freunberger, S. A., Peng, Z. & Bruce, P. G. The carbon electrode in nonaqueous Li-O₂ cells. *J. Am. Chem. Soc.* **135**, 494–500 (2013).
30. Bikkarolla, S. K., Cumson, P., Joseph, P. & Papakonstantinou, P. Oxygen reduction reaction in electrochemically reduced graphene oxide. *Faraday Discuss.* **173**, 415–428 (2014).
31. Antolini, E. Iridium as catalyst and cocatalyst for oxygen evolution/reduction in acidic polymer electrolyte membrane electrolyzers and fuel cells. *ACS Catal.* **4**, 1426–1440 (2014).
32. Donkersloot, H. C. & Van Vucht, J. H. N. The crystal structure of IrLi, Ir₃Li and LiRh₃. *J. Less Common Met.* **50**, 279–282 (1976).
33. Pouget, E. M. *et al.* The initial stages of template-controlled CaCO₃ formation revealed by cryo-TEM. *Science* **323**, 1455–1458 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was primarily supported by the US Department of Energy under contract DE-AC02-06CH11357 from the Vehicle Technologies Office, Department of Energy, Office of Energy Efficiency and Renewable Energy. We also acknowledge support from the Center for Electrochemical Energy Science (CEES), an Energy Frontier Research Center (EFRC) funded by the US Department of Energy, Office of Science, Office of Basic Energy Sciences (X-ray measurements and analysis). We also acknowledge support from the University of Illinois-Chicago Chancellor Proof of Concept Fund (DEMS measurements). We acknowledge the Conn Renewable Energy Research Center at the University of Louisville for providing the access to the DEMS equipment. We acknowledge grants of computer time through INCITE awards on the BlueGene/Q computer at Argonne National Laboratory and allocations on the CNM Carbon Cluster at Argonne National Laboratory and the LCRC Fusion Cluster at Argonne National Laboratory. Use of the Advanced Photon Source and the Electron Microscopy Center, Center for Nanoscale Materials was supported by the US Department of Energy, Office of Basic Energy Sciences, under contract no. DE-AC02-06CH11357. We acknowledge financial support from the Human Resources Development of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) funded by the Korea government Ministry of Knowledge Economy (no. 20124010203310), and from the Basic Science Research Program (no. NRF-2014R1A2A1A11049801). We acknowledge C. Barile, R. Rooney, R. Assary and P. Redfern for discussions and help on the lithium superoxide reaction mechanism.

Author Contributions J.L. and K.A. designed the experiments; Y.J.L., J.-B.P. and Y.S.J. synthesized the cathode materials; J.L., D.J.M. and J.W. performed and analysed the TEM imaging experiments; J.L., X.L., L.A.C. and Z.C. performed and analysed the X-ray measurements; J.L., X.L., Z.Z.F., D.Z. and H.-H.W. tested the cathode materials; M.A., A.S.-K. and B.K. performed the DEMS measurements, H.-H.W., X.L. and S.B. performed the Raman, NMR, EPR and FTIR experiments, K.C.L. and L.A.C. were responsible for the theoretical computations. L.A.C., K.A. and Y.-K.S. supervised the project; L.A.C., J.L. and K.A. wrote the paper. All of the authors discussed the results and reviewed the manuscript.

Author Information Atomic coordinates for the LiO₂ crystal structure from DFT can be obtained from the ICSD Database (http://www2.fiz-karlsruhe.de/icsd_home.html). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.A.C. (curtiss@anl.gov) or K.A. (amine@anl.gov) or Y.-K.S. (yksun@hanyang.ac.kr).

METHODS

Material preparation. Graphene oxide (GO) was prepared by a modified Hummers method^{21,22}. Pristine reduced GO (rGO) was produced by reducing GO in polyol³⁴. Specifically, a GO dispersion (1 mg ml⁻¹) in ethylene glycol (EG) was first prepared with the aid of horn sonication for 1 h. The solution pH was then adjusted to pH 13 with NaOH (2.5 M in EG). The reaction temperature was increased to 120 °C, and a reducing agent (NaBH₄) dissolved in EG was injected slowly. The solution was allowed to react for 1 h at this temperature and then cooled to room temperature. The precipitate was filtered, washed and dried under vacuum. The Ir-rGO composite was synthesized by a hydrothermal method. IrCl₃·H₂O was added to 100 ml aqueous GO solution (0.67 mg ml⁻¹) with a weight ratio of IrCl₃·H₂O to GO of 1.9:1. The mixture was stirred for 2 h. The resulting solution was transferred to a Teflon-lined autoclave and there reacted hydrothermally at 180 °C for 12 h. The precipitate was filtered, washed and dried under vacuum. During hydrothermal treatment, Ir nanocrystals are obtained with simultaneous reduction of GO into rGO, without addition of strong base.

After synthesis, the Ir-rGO and rGO powders were dried at 80 °C under vacuum for 24 h. To prepare the oxygen electrode, the Ir-rGO and rGO were intimately mixed in an N-methyl-2-pyrrolidone (NMP) liquid and a polyvinylidene fluoride binder (PVDF) with a weight ratio of active ingredients to PVDF of 8:2. The slurry was coated onto a gas-diffusion layer (TGP-H-030 carbon paper, Torray) and dried for 12 h at 100 °C under vacuum to remove the residual solvent.

Electrochemical testing. Electrochemical testing was carried out using a Swagelok-type cell composed of a lithium metal anode, electrolyte (1 M LiCF₃SO₃ in tetraethylene glycol dimethyl ether (TEGDME) impregnated into a glass fibre separator), and a porous cathode (7/16 inch diameter). The cells were sealed except for the Al grid window that exposed the porous cathode to 1 bar O₂ pressure. The electrochemical measurements were carried out using a MACCOR cyclor. The discharge-charge performance was conducted in the voltage range of 2.2–4.5 V at a constant current density of 100 mA g⁻¹, and the cell was maintained in 1 bar O₂ atmosphere to avoid any negative effects of humidity and CO₂.

Characterization techniques. The phase structures of the discharge products were identified using high-energy X-ray diffraction (HE-XRD) with a wavelength of 0.11165 Å, performed at beamline 11ID-C of Sector 11 at the Advanced Photon Source (APS) of Argonne National Laboratory. The X-ray specimens were sealed with Kapton tape as a protective film in the glove box to avoid any side reactions from the air. The HE-XRD patterns were collected in the transmission mode. During the course of the measurements, a high-energy X-ray beam hit the sample horizontally, and a 2D detector (Perkin Elmer large area detector) was used to collect the X-ray diffraction profiles using transmission mode. The 2D patterns were then integrated into conventional 1D patterns (intensity versus 2θ) for final data analysis using Fit2d software.

Scanning transmission electron microscopy (TEM; JEOL JEM-2100F FEG FasTEM with an accelerating voltage of 80 kV) was employed to evaluate the morphology and particle size of the Ir catalysts and the discharge products on the porous cathodes. Spherical and chromatic aberration correction enables the microscope to reach the information limit, which is a resolution of 0.1 nm (measured by Young's fringes) at 80 kV. To prepare the TEM specimens, a dilute suspension was prepared by ultrasonically dispersing the samples in dimethyl ether for 5 min, and a drop of the suspension was placed onto a copper grid and dried. Particle size histograms were generated from the TEM images using software ImageJ. Field-emission scanning electron microscopy (SEM, Hitachi S-4700) coupled with backscattering electron imaging (BSE) was employed to determine the morphology and estimate the particle size of Ir catalyst and discharge products.

Raman spectra of the discharged cathode were obtained using a Renishaw 2000 or inVia microscope spectrometer with a HeNe laser at an exciting wavelength of 633 nm. The sample was loaded inside a glove box into a gas-tight Raman cell with a glass or quartz window. Raman spectrum collection was set up in a 180° reflective mode. Roughly 10% of the maximum 13 mW laser intensity was applied. Collection time constant setting varied from 30 s to ~100 s.

Nuclear Magnetic Resonance (NMR) spectra were collected on a Bruker Avance III 500 MHz (11.7 T) spectrometer in deuterated tetrahydrofuran (THF-d₈) or deuterium oxide (D₂O). Battery samples were cycled the appropriate number of times before disassembly and rinsing with a deuterated solvent. For the samples in THF-d₈, NMR samples were prepared entirely under an inert atmosphere. For samples in D₂O, the battery was disassembled under an inert atmosphere, removed from the glove box, immediately rinsed with D₂O, and sealed in an NMR tube. All 1H-NMR collections were sampled for 128 scans, manually phased, and baseline corrected before peak integration. Data collection was performed using the Bruker Topspin software (v. 3.1). Data analysis used either TopSpin (v. 3.1) or SpinWorks (V4.1.0.0).

DEMS measurements were carried out to examine the type of discharge product by measuring the evolved oxygen (O₂) molecules during the charging process, as well as calculating the e⁻/O₂ ratio. The experiments were performed using an HPR40 (Hiden Analytical) instrument directly connected to a customized Swagelok battery set-up.

Theoretical methods. To study the stability of the LiO₂ system (that is, crystal, crystalline surfaces, amorphous-like thin film), its interface with a solvent and its surface growth on an intermetallic substrate (that is, Ir₃Li), we carried out Density Functional Theory (DFT) calculations with plane wave basis sets as implemented in the VASP code^{35,36}. All the calculations were spin-polarized and carried out using the gradient corrected exchange-correlation functional of Perdew, Burke and Ernzerhof (PBE)³⁷ under the projector augmented wave (PAW) method, with plane wave basis sets up to a kinetic energy cutoff of 400 eV. The PAW method³⁸ was used to represent the interaction between the core and valence electrons, and the Kohn-Sham valence states (that is, 1s for H, 2s for Li, 2s 2p for C and O, 6s 5d for Ir) are expanded in plane wave basis sets. For the geometry optimization and Nudge Elastic Band calculations, the convergence criterion of the total energy was set to be within 1 × 10⁻⁵ eV for the k-point integration, and all the atoms and geometries were optimized until the residual forces became less than 1 × 10⁻² eV Å⁻¹. For the LiO₂ crystal, the calculation was based on a mesh of 9 × 9 × 9 in k-point grid. For both the crystalline and amorphous-like LiO₂ thin film surfaces, a k-point grid of 6 × 6 × 1 was used. For the *Ab initio* Molecular Dynamics (AIMD) simulations, all the calculations were carried out with plane wave basis sets up to a kinetic energy cutoff of 300 eV. For the simulations of LiO₂ surface with some solvent molecules on it, the Van der Waals method of Grimme (that is, DFT-D2)³⁹ was used throughout both the DFT and AIMD calculations. For the simulation of the solvent, a smaller ether solvent molecule, that is, dimethoxyethane (DME) was used instead of tetraethylene glycol dimethyl ether (TEGDME) in order to reduce the computational cost. To investigate the thermodynamic stability of the system at room temperature, all the structures from the DFT optimizations were then thermally equilibrated at T = 300 K using AIMD simulations based on an Nose-Hoover NVT-ensemble with a time step of 1 fs.

For the *ab initio* prediction of Electron Paramagnetic Resonance (EPR) g-tensors of the crystalline solids LiO₂ and NaO₂, we have used the DFT plane wave code Quantum Espresso simulation package⁴⁰ in which the Gauge-Including Projector Augmented Plane Wave (GIPAW) approach⁴¹ has been implemented. To be consistent with our VASP calculations, we employed the PBE functional and Martins-Troullier norm-conserving pseudopotentials generated by D Ceresoli (<https://sites.google.com/site/dceresoli/pseudopotentials>). Throughout the simulation, a plane wave energy cutoff of 150 Ry and k-point mesh (9 × 9 × 9) was used. Only the NaO₂ solid in *Pnmm* orthorhombic phase⁴² that is thermodynamically stable at low temperature was considered in our DFT EPR calculations. For the *ab initio* prediction of the EPR g-tensor of the LiO₂ and NaO₂ molecules, all the calculations were obtained using the Gaussian09 code. The accuracy of the theoretical predictions are systematically tested using PBE and B3LYP functionals with the 6-31+G(2df,p) and aug-cc-PVQZ basis sets for direct comparison to the reported experimental LiO₂ and NaO₂ EPR g-tensors^{43,44}.

Sample size. No statistical methods were used to predetermine sample size.

34. Ha, H.-W., Kim, I. Y., Hwang, S.-J. & Ruoff, R. S. One-pot synthesis of platinum nanoparticles embedded on reduced graphene oxide for oxygen reduction in methanol fuel cells. *Electrochem. Solid-State Lett.* **14**, B70–B73 (2011).
35. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
36. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
37. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
38. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
39. Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **27**, 1787–1799 (2006).
40. Giannozzi, P. et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter* **21**, 395502 (2009).
41. Pickard, C. J. & Mauri, F. First-principles theory of the EPR g tensor in solids: defects in quartz. *Phys. Rev. Lett.* **88**, 086403 (2002).
42. Ziegler, M., Rosenfeld, M., Kaenzig, W. & Fischer, P. Strukturuntersuchungen an Alkalihyperoxiden. *Helvetica Physica Acta* **49**, 57–90 (1976).
43. Adrian, F. J., Cochran, E. L. & Bowers, V. A. ESR spectra and structures of NaO₂ and NaO₂. *J. Chem. Phys.* **59**, 56–60 (1973).
44. Lindsay, D. M. & Garland, D. A. ESR spectra of matrix-isolated lithium superoxide. *J. Phys. Chem.* **91**, 6158–6161 (1987).

Future ocean hypercapnia driven by anthropogenic amplification of the natural CO₂ cycle

Ben I. McNeil^{1,2} & Tristan P. Sasse^{1,3}

High carbon dioxide (CO₂) concentrations in sea-water (ocean hypercapnia) can induce neurological, physiological and behavioural deficiencies in marine animals^{1–10}. Prediction of the onset and evolution of hypercapnia in the ocean requires a good understanding of annual variations in oceanic CO₂ concentration, but there is a lack of relevant global observational data. Here we identify global ocean patterns of monthly variability in carbon concentration using observations that allow us to examine the evolution of surface-ocean CO₂ levels over the entire annual cycle under increasing atmospheric CO₂ concentrations. We predict that the present-day amplitude of the natural oscillations in oceanic CO₂ concentration will be amplified by up to tenfold in some regions by 2100, if atmospheric CO₂ concentrations continue to rise throughout this century (according to the RCP8.5 scenario of the Intergovernmental Panel on Climate Change)¹¹. The findings from our data are broadly consistent with projections from Earth system climate models^{12–15}. Our predicted amplification of the annual CO₂ cycle displays distinct global patterns that may expose major fisheries in the Southern, Pacific and North Atlantic oceans to hypercapnia many decades earlier than is expected from average atmospheric CO₂ concentrations. We suggest that these ocean ‘CO₂ hotspots’ evolve as a combination of the strong seasonal dynamics of CO₂ concentration and the long-term effective storage of anthropogenic CO₂ in the oceans that lowers the buffer capacity in these regions, causing a nonlinear amplification of CO₂ concentration over the annual cycle. The onset of ocean hypercapnia (when the partial pressure of CO₂ in sea-water exceeds 1,000 micro-atmospheres) is forecast for atmospheric CO₂ concentrations that exceed 650 parts per million, with hypercapnia expected in up to half the surface ocean by 2100, assuming a high-emissions scenario (RCP8.5)¹¹. Such extensive ocean hypercapnia has detrimental implications for fisheries during the twenty-first century.

It has been shown that even relatively moderate ocean hypercapnia, corresponding to an approximate doubling of present-day oceanic levels of CO₂, is detrimental to a range of tropical and temperate fish species, with experiments showing substantial neurological, behavioural and physiological effects when fish are exposed to short periods (4–40 days) during which the partial pressure of CO₂ in sea-water (p_{CO_2}) is 550–1,000 μatm (refs 1–7). Similar effects of hypercapnia have also been reported for some marine invertebrates^{9,10}. High-CO₂ conditions may also affect receptors for GABA (γ -aminobutyric acid), the main inhibitory neurotransmitter in the vertebrate brain⁸. Given the prevalence and importance of GABA_A receptors in animals including fish, behavioural changes induced by high-CO₂ conditions could affect numerous fish species and have wide-reaching implications for population replenishment, community structure and ecosystem function^{5,8}. Predicting the onset and evolution of hypercapnia requires an understanding of CO₂ variability in the ocean over the entire annual cycle, something that is currently lacking as a result of sparse global carbon observations. In particular, the natural variation in ocean inorganic

carbon state variables (total dissolved inorganic carbon (C_T) and total alkalinity (A_T)) is poorly understood, but is necessary to project future CO₂ concentrations in the ocean.

Here we identify and exploit an observationally derived monthly climatology for inorganic carbon state variables (C_T and A_T) throughout the surface ocean for the nominal year of 2000 (ref. 16). We use this data as a baseline to project the likely future evolution of monthly p_{CO_2} levels to the year 2100 under a Representative Concentration Pathway 8.5 (RCP8.5)¹¹ scenario in the atmosphere. Our data-based projections assume a constant air–sea CO₂ disequilibrium and a steady-state seasonal cycle of C_T and A_T (see Methods). To explore how variations in air–sea CO₂ gas exchange, future warming and other biogeochemistry may alter our projections, we compare our results with output from four different Earth system models (ESMs; see Methods).

We find a substantial amplification of the annual oceanic CO₂ cycle over the twenty-first century (Fig. 1). In particular, for regions within the Southern, Pacific and North Atlantic oceans, our data-based projections show a five- to eightfold amplification of CO₂ concentration in the annual CO₂ cycle over the twenty-first century under increasing atmospheric CO₂ concentration (Fig. 1c). ESMs suggest an even greater CO₂ amplification in many parts of the ocean (Fig. 1d, Extended Data Fig. 3). For example, the UK Met Office Hadley Centre climate model predicts that the natural CO₂ cycle becomes amplified by eight- to tenfold throughout large parts of both the Southern and Arctic oceans (Fig. 1d). Other models predict up to tenfold CO₂ amplification in regions including the eastern Pacific, Antarctic, Arctic and North Atlantic oceans (Extended Data Fig. 3).

The geographic extent and location of the largest future CO₂ amplification varies, but both our data-based projections and models generally show that the Southern, equatorial Pacific and North Atlantic oceans are hotspots for future CO₂ amplification. The largest regional difference between our data-based projections and models is within the Arctic Ocean, for which most models predict a five- to tenfold future CO₂ amplification (see Extended Data Fig. 3), whereas our data-based approach does not (Fig. 1). However, there is very little seasonal CO₂ data relating to the Arctic Ocean, so it is likely that our data-based approach does not capture variability well in this region. Furthermore, our data-based projections show a greater future amplification (approximately threefold) throughout most of the ocean in comparison to the models (approximately twofold) (Fig. 1)—a consequence of varying caveats and assumptions between both approaches (see Methods). Despite this, there is good agreement on broad zones of intense CO₂ amplification (five- to tenfold) between our data-based approach and models within the Southern, equatorial Pacific and North Atlantic oceans (Extended Data Fig. 3).

The substantial future CO₂ amplification is projected to cause a large change in the distribution of p_{CO_2} throughout the ocean, with organisms having to experience much larger variability and extremes in surface-ocean p_{CO_2} levels throughout the twenty-first century (Fig. 2). Although this amplification causes more CO₂ variability in the future, it is strongly nonlinear, with greater amplification for higher p_{CO_2} levels,

¹Climate Change Research Centre, University of New South Wales, Sydney, New South Wales, Australia. ²Thinkable.org, 299 Sussex Street, Sydney, New South Wales, Australia.

³School of Mathematics and Statistics, University of New South Wales, Sydney, New South Wales, Australia.

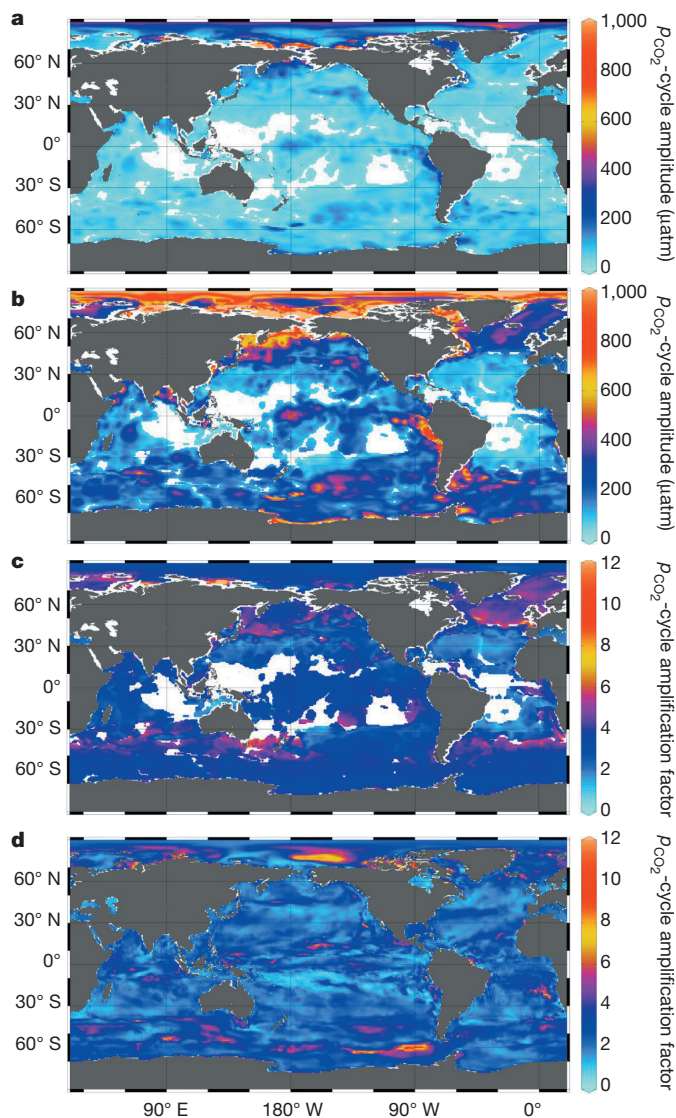


Figure 1 | Future amplification of ocean CO₂ levels. **a, b,** The amplitude (difference between monthly maximum and minimum) of the annual p_{CO_2} cycle for 2000 (**a**) and 2100 (**b**). White regions here and elsewhere correspond to those where the amplitude of the annual p_{CO_2} cycle in the year 2000 is lower than the p_{CO_2} uncertainty of our SOMLO approach ($< 27 \mu\text{atm}$). **c,** Data-based projection of the p_{CO_2} amplification factor in the surface ocean by the year 2100 (see Methods). The p_{CO_2} amplification factor is calculated by dividing the amplitude of the annual p_{CO_2} cycle in 2100 (shown in **b**) by that in 2000 (shown in **a**). A factor of zero implies that the annual cycle of p_{CO_2} is not projected to change by 2100; a factor of ten implies that the amplitude of the annual cycle is projected to increase by an order of magnitude by 2100. **d,** The p_{CO_2} amplification factor in the surface ocean by 2100 as predicted from the HadGEM2-ES global climate model¹⁴. (HadGEM2 is a coupled ESM that was used by the UK Met Office Hadley Centre for the CMIP-5 climate model intercomparison project³¹.) These and the other maps were generated using Ocean Data View (<http://odv.awi.de/>).

thereby accelerating the onset of extreme CO₂ conditions (see Methods). The SOMLO (self-organizing multiple linear output¹⁶; see Methods) data-based approach tends to predict higher future p_{CO_2} extremes than do models (Fig. 2). Upon further analysis, this difference is due to a combination of models under-estimating observed natural variability, and our steady-state assumptions, which probably overestimate future p_{CO_2} concentrations in some regions by about 10% when using the SOMLO approach (see Methods for discussion).

To illustrate the generalized, nonlinear, p_{CO_2} future amplification we find here, we plot results for a nominal region in the subpolar North

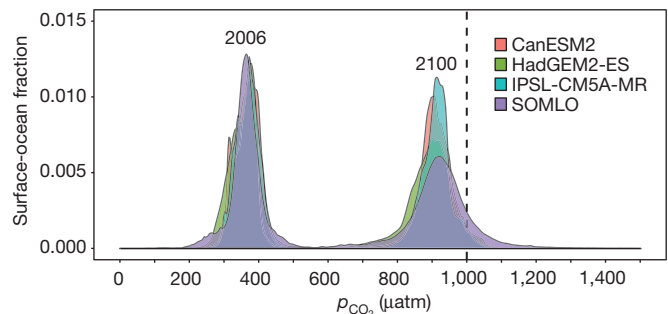


Figure 2 | Increasing variability in future CO₂ levels. Surface-ocean distribution of monthly p_{CO_2} levels in 2006 and in 2100 as predicted from three different ESMs (CanESM2, HadGEM2-ES and IPSL-CM5A-MR; see Methods for details) and our data (SOMLO). The vertical dashed line at $p_{\text{CO}_2} = 1,000 \mu\text{atm}$ indicates the onset of hypercapnia.

Pacific Ocean (Fig. 3). On the basis of our data-based estimates for the year 2000, the amplitudes of the annual cycle in C_T and A_T were determined to be $91 \mu\text{mol kg}^{-1}$ and $19 \mu\text{mol kg}^{-1}$, respectively; these amplitudes are controlled by a combination of biological processes, ocean mixing and air–sea gas exchange. For a given time-varying temperature and salinity over the annual cycle, these inorganic carbon conditions translate to a p_{CO_2} annual cycle of about $90 \mu\text{atm}$ (Fig. 3). Assuming the natural cycle of C_T and A_T remains constant in the future, we predict that the magnitude of the annual p_{CO_2} cycle will increase to over $400 \mu\text{atm}$ by 2100—a more than fourfold amplification (Fig. 3). A consequence of this amplification is that critical high- p_{CO_2} thresholds are passed earlier than is expected from atmospheric CO₂ concentrations alone (Fig. 3), exposing marine organisms to hypercapnia many decades earlier than previously anticipated.

We now consider the causes of the nonlinear amplification of the natural CO₂ cycle. The total dissolved inorganic carbon concentration in the ocean (C_T) can be generalized as the sum of three species: dissolved carbon dioxide (p_{CO_2}), carbonate ions (CO_3^{2-}) and bicarbonate ions (HCO_3^-). Bicarbonate ions makes up around 90% of the total inorganic carbon pool; p_{CO_2} is a very minor component at about 0.5% of the total. As the total inorganic carbon pool is used and undergoes changes via biological processes or mixing, instantaneous chemical equilibrium in the carbonate system in sea-water induces a coincident, but variable, change in p_{CO_2} that is defined by the Revelle factor¹⁷. The Revelle factor (R), sometimes called the ‘buffer factor’, describes the relationship between the relative change in p_{CO_2} that is induced by a change in total inorganic carbon:

$$R = \frac{\partial \ln(p_{\text{CO}_2})}{\partial \ln(C_T)}$$

As R changes in space and time, so too does p_{CO_2} for a given change in C_T , as a result of natural variability.

We demonstrate how the Revelle factor mediates p_{CO_2} conditions using a hypothetical situation in which organic matter respiration increases C_T by $100 \mu\text{mol kg}^{-1}$ in two different oceanic regions. The Southern Ocean with low carbonate ion concentration, for example, maintains a higher surface Revelle factor of $R = 16$ than does the subtropics, where $R \approx 8$ (see Methods). An increase of C_T by $100 \mu\text{mol kg}^{-1}$ due to organic matter respiration would increase p_{CO_2} by approximately $163 \mu\text{atm}$ in the subtropics, but approximately $521 \mu\text{atm}$ in the Southern Ocean. Although the Southern Ocean has a Revelle factor that is double that of the subtropics, the resultant change in p_{CO_2} is more than triple that of the subtropics, owing to the nonlinear carbon chemistry¹⁷. A larger Revelle factor induces a much greater p_{CO_2} response for a given biogeochemical change in total inorganic carbon (see Methods).

Our finding of a substantial amplification of future CO₂ is in part due to long-term anthropogenic changes to Revelle factors throughout the surface ocean. As anthropogenic CO₂ is absorbed by the ocean, the

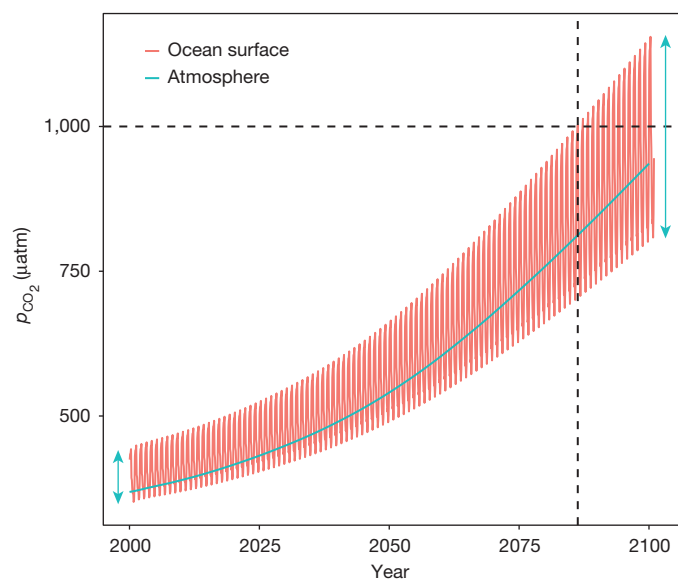


Figure 3 | Example of future CO₂ amplification and early-onset hypercapnia. Monthly p_{CO_2} cycle over the twenty-first century in a nominal region in the North Pacific surface ocean (pink) illustrating the early onset of monthly oceanic hypercapnia events in comparison to atmospheric CO₂ (blue). The dashed horizontal line indicates the onset of hypercapnia at $p_{\text{CO}_2} = 1,000 \mu\text{atm}$; the dashed vertical line marks the time at which the surface ocean first experiences hypercapnia events. The fourfold amplification of the annual p_{CO_2} cycle between 2000 and 2100 that causes the early onset of monthly hypercapnia events is highlighted by the blue arrows.

carbonate ion concentration decreases¹⁸, resulting in an observed long-term increase of about one to Revelle factors throughout the ocean¹⁹. For subtropical waters, Revelle factors are projected to increase from about 9 to 13 by 2100 (ref. 20). These decadal-scale anthropogenic increases in Revelle factors have the effect of amplifying natural CO₂ variability over the annual cycle, causing the strongly nonlinear future amplification of the natural CO₂ cycle²⁰ that we observe over most of the ocean.

We examine how future CO₂ amplification alters the likely onset, patterns and dynamics of month-long hypercapnia in marine organisms under increasing atmospheric CO₂ levels. The CO₂ level leading to the onset of hypercapnia varies for different marine organisms; however, we assume that hypercapnia in marine organisms is coincident with hypercapnia in sea-water, which we define to be when p_{CO_2} exceeds 1,000 μatm for a month-long period. This assumption is justified by observations from recent laboratory studies in temperate fish, and is consistent with other studies^{21–23}. However, the effect of future nonlinear amplification and high-CO₂ hotspots that we find here occurs irrespective of the hypercapnia threshold level chosen.

We find that surface-ocean hypercapnia ($p_{\text{CO}_2} > 1,000 \mu\text{atm}$) in large parts of the ocean probably begins when atmospheric CO₂ reaches about 650 p.p.m., which occurs around the year 2060 under RCP8.5 (Fig. 4a). Hypercapnia starts in ‘hotspots’ within the eastern Pacific, Southern and North Pacific oceans, but expands to cover large areas of all ocean basins by about 2080. The onset and exposure time of hypercapnia is highly dependent on region, with large parts of the Southern, Pacific and North Atlantic oceans likely to experience 5–12 months of hypercapnia by the end of the twenty-first century even though atmospheric CO₂ is predicted to remain well below 1,000 p.p.m. (Fig. 4b). Importantly, our results suggest that hypercapnia occurs earliest, and with most severity, in major fishery zones that make up a large proportion of the 80 million tonnes of commercial fish caught worldwide per year (ref. 24). There are more than 23,000 Teleost fish species, which make up the majority of the important sport and commercial fish species²⁵. Our results suggest

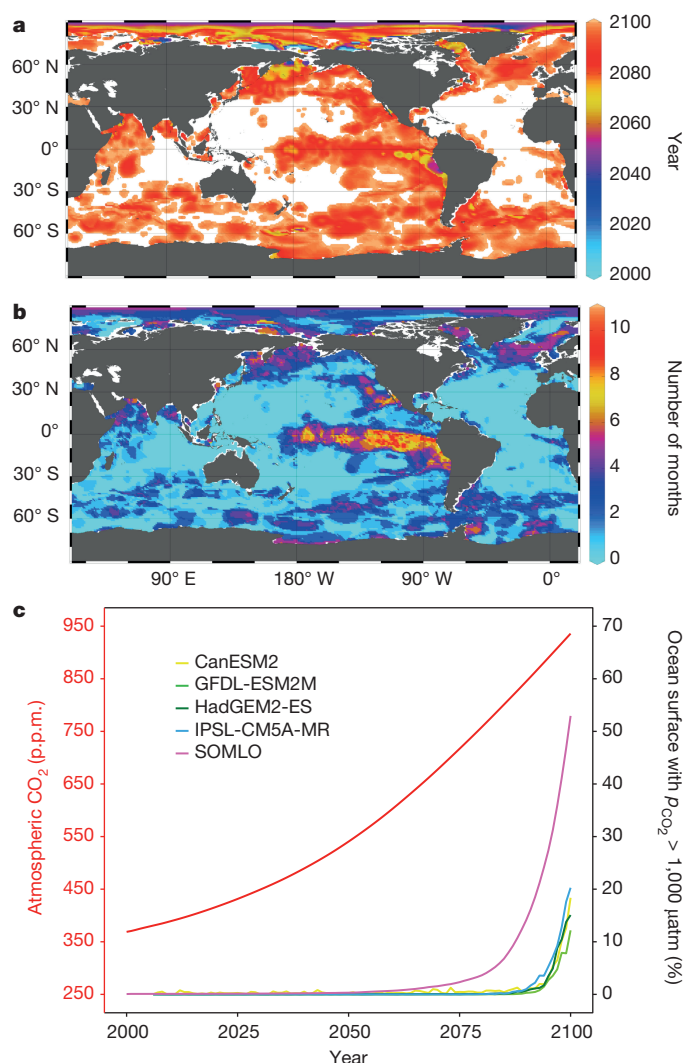


Figure 4 | Global onset of future hypercapnia. **a**, The year in which monthly hypercapnia events begin ($p_{\text{CO}_2} > 1,000 \mu\text{atm}$) assuming the RCP8.5 atmospheric CO₂ scenario, determined from our data-based projections. **b**, Number of months in 2100 in which the surface-ocean p_{CO_2} levels exceed 1,000 μatm assuming the RCP8.5 atmospheric CO₂ scenario, determined from our data-based projections. **c**, Geographic extent of month-long hypercapnia events in the surface ocean throughout the twenty-first century. Atmospheric CO₂ levels (measured by the left axis) are shown in red. The fraction of the surface ocean that experiences month-long hypercapnia events (measured by the right axis) as determined by the SOMLO data-based projection (pink) and four climate models (yellow, green, blue and purple, as indicated).

the potential for substantial implications for future commercial fisheries, particularly if the threshold for month-long hypercapnia in Teleost fish species is less than 1,000 μatm , as some studies^{1–7} have found.

By the end of the twenty-first century, month-long hypercapnia events are projected to occur in up to half the surface ocean under a high-emission atmospheric CO₂ scenario (Fig. 4c). Models suggest the onset of hypercapnia when atmospheric CO₂ reaches about 750 p.p.m. assuming the high-emissions (RCP8.5) scenario, with up to 22% of the surface ocean exhibiting hypercapnia by the year 2100 (Fig. 4c). These model predictions of future hypercapnia are probably a substantial underestimation; our analysis and other studies^{26–29} suggest that these models under-represent observed monthly CO₂ variability (see Methods). To better predict ocean hypercapnia in the future, it is therefore critical that data and models are consistent in terms of constraining high-frequency CO₂ variability (days to months) in the ocean. The amplification of the natural oceanic CO₂ cycle we find here under

RCP8.5 will probably scale in proportion to variations in future atmospheric CO₂ increases. Future hypercapnia and its impact on the ocean will therefore need to be explored under a range of emission scenarios.

Coastal and intertidal marine environments naturally experience large CO₂ fluctuations³⁰. However long-term anthropogenic-induced changes to carbon chemistry are likely to induce a substantial nonlinear amplification of future CO₂ variability both within the coastal ocean²⁰ and the wider open ocean, as we show here. Many marine organisms have naturally evolved to be resilient to large fluctuations in CO₂ conditions in coastal environments. Marine organisms within the open ocean, however, are not accustomed to extreme fluctuations in CO₂. The regions that are likely to experience both early onset of hypercapnia and extreme fluctuations in p_{CO_2} also host substantial commercial fisheries. The appearance of month-long hypercapnia events may have considerable implications for fisheries during the twenty-first century.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 31 March; accepted 13 October 2015.

- Cripps, I. L., Munday, P. L. & McCormick, M. I. Ocean acidification affects prey detection by a predatory reef fish. *PLoS ONE* **6**, e22736 (2011).
- Devine, B. M. & Munday, P. L. Habitat preferences of coral-associated fishes are altered by short-term exposure to elevated CO₂. *Mar. Biol.* **160**, 1955–1962 (2013).
- Devine, B. M., Munday, P. L. & Jones, G. P. Homing ability of adult cardinalfish is affected by elevated carbon dioxide. *Oecologia* **168**, 269–276 (2012).
- Dixon, D. L., Munday, P. L. & Jones, G. P. Ocean acidification disrupts the innate ability of fish to detect predator olfactory cues. *Ecol. Lett.* **13**, 68–75 (2010).
- Ferrari, M. C. O. *et al.* Putting prey and predator into the CO₂ equation – qualitative and quantitative effects of ocean acidification on predator–prey interactions. *Ecol. Lett.* **14**, 1143–1148 (2011).
- Jutfelt, F., De Souza, K. B., Vuylsteke, A. & Sturve, J. Behavioural disturbances in a temperate fish exposed to sustained high-CO₂ levels. *PLoS ONE* **8**, e65825 (2013).
- Munday, P. L. *et al.* Ocean acidification impairs olfactory discrimination and homing ability of a marine fish. *Proc. Natl Acad. Sci. USA* **106**, 1848–1852 (2009).
- Nilsson, G. E. *et al.* Near-future carbon dioxide levels alter fish behaviour by interfering with neurotransmitter function. *Nature Clim. Change* **2**, 201–204 (2012).
- Spady, B. L., Watson, S. A., Chase, T. J. & Munday, P. L. Projected near-future CO₂ levels increase activity and alter defensive behaviours in the tropical squid *Idiosepius pygmaeus*. *Biol. Open* **3**, 1063–1070 (2014).
- Watson, S. A. *et al.* Marine mollusc predator-escape behaviour altered by near-future carbon dioxide levels. *Proc. R. Soc. Lond. B* **281**, 20132377 (2014).
- Meinshausen, M. *et al.* The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Clim. Change* **109**, 213–241 (2011).
- Aumont, O. & Bopp, L. Globalizing results from ocean in situ iron fertilization studies. *Glob. Biogeochem. Cycles* **20**, GB2017 (2006).
- Dunne, J. P. *et al.* GFDL's ESM2 global coupled climate–carbon earth system models. Part II: carbon system formulation and baseline simulation characteristics. *J. Clim.* **26**, 2247–2267 (2013).
- Jones, C. D. *et al.* The HadGEM2-ES implementation of CMIP5 centennial simulations. *Geosci. Model Dev.* **4**, 543–570 (2011).
- Yang, D. & Saenko, O. A. Ocean heat transport and its projected change in CanESM2. *J. Clim.* **25**, 8148–8163 (2012).
- Sasse, T. P., McNeil, B. I. & Abramowitz, G. A novel method for diagnosing seasonal to inter-annual surface ocean carbon dynamics from bottle data using neural networks. *Biogeosciences* **10**, 4319–4340 (2013).
- Egleston, E. S., Sabine, C. L. & Morel, F. M. M. Revelle revisited: buffer factors that quantify the response of ocean chemistry to changes in DIC and alkalinity. *Glob. Biogeochem. Cycles* **24**, GB1002 (2010).
- Orr, J. C. *et al.* Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. *Nature* **437**, 681–686 (2005).
- Sabine, C. L. *et al.* The oceanic sink for anthropogenic CO₂. *Science* **305**, 367–371 (2004).
- Shaw, E. C., McNeil, B. I., Tilbrook, B., Matear, R. & Bates, M. L. Anthropogenic changes to seawater buffer capacity combined with natural reef metabolism induce extreme future coral reef CO₂ conditions. *Glob. Change Biol.* **19**, 1632–1641 (2013).
- Byrne, M., Lamare, M., Winter, D., Dworjanyan, S. A. & Uthicke, S. The stunting effect of a high CO₂ ocean on calcification and development in sea urchin larvae, a synthesis from the tropics to the poles. *Phil. Trans. R. Soc. Lond. B* **368**, 20120439 (2013).
- Fabry, V. J., Seibel, B. A., Feely, R. & Orr, J. C. Impacts of ocean acidification on marine fauna and ecosystem processes. *ICES J. Mar. Sci.* **65**, 414–432 (2008).
- Maas, A. E., Wishner, K. F. & Seibel, B. A. The metabolic response of pteropods to acidification reflects natural CO₂-exposure in oxygen minimum zones. *Biogeosciences* **9**, 747–757 (2012).
- Food and Agriculture Organization of the United Nations. *The State of World Fisheries and Aquaculture 2014* 9–12 (FAO, 2014).
- Volf, J.-N. Genome evolution and biodiversity in teleost fish. *Heredity* **94**, 280–294 (2013).
- Ishii, M. *et al.* Air–sea CO₂ flux in the Pacific Ocean for the period 1990–2009. *Biogeosciences* **11**, 709–734 (2014).
- Lenton, A. *et al.* Sea–air CO₂ fluxes in the Southern Ocean for the period 1990–2009. *Biogeosciences* **10**, 4037–4054 (2013).
- Sarma, V. V. S. S. *et al.* Sea–air CO₂ fluxes in the Indian Ocean between 1990 and 2009. *Biogeosciences* **10**, 7035–7052 (2013).
- Schuster, U. *et al.* Trends in North Atlantic sea-surface fCO₂ from 1990 to 2006. *Deep Sea Res. II* **56**, 620–629 (2009).
- Shaw, E. C., McNeil, B. I. & Tilbrook, B. Impacts of ocean acidification in naturally variable coral reef flat ecosystems. *J. Geophys. Res. Oceans* **117**, C03038 (2012).
- Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).

Acknowledgements We thank R. Matear and A. Lenton from the CSIRO for conversations, performing initial model runs and analysis of CMIP5 model output for use here. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups (Hadley Centre, IPSL, Can and GFDL) for producing their models and making the output openly accessible. For CMIP, the US Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support, and led the development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We acknowledge support from the Australian Research Council discovery grant DP110104955.

Author Contributions B.I.M. conceived and designed the project, interpreted the results and wrote the paper. T.P.S. helped design the project and carried out all of the numerical and data analysis, and contributed important interpretations, discussions and revisions to the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.I.M. (b.mcneil@unsw.edu.au).

METHODS

Data. No statistical methods were used to predetermine sample size.

For over twenty years global oceanographic measurement programs like the World Ocean Circulation Experiment (WOCE) and Climate Variability and Predictability (CLIVAR) have collected and analysed hundreds of thousands of *in situ* bottle (discrete) carbon measurements of total dissolved carbon dioxide (C_T) and total alkalinity (A_T) along with standard hydrographic parameters (SHPs; temperature, salinity, dissolved oxygen and nutrients). For this study we use the updated global surface-ocean *in situ* concentrations of C_T and A_T (ref. 32).

SOMLO approach to quantify present-day carbon variability. Although spatiotemporal coverage of the global bottle-derived C_T and A_T dataset is too sparse on its own, the corresponding SHPs provide biogeochemical information that offers an independent constraint on C_T and A_T where only SHP data exist. For this study we exploit this SHP data by coupling a neural-network-clustering algorithm with a principle-component regression to identify monthly surface C_T and A_T distributions from SHP climatologies. In our approach, the algorithm captures larger-scale ocean dynamics by clustering data into 'biogeochemical fingerprints' in a self-organizing map (SOM)³³. In brief, the SOM approach uses bottle-derived C_T and A_T measurements and SHP distribution information, along with geographical constraints, to iteratively cluster the bottle measurements into a set of neurons on the basis of similarities and homogeneity within the dataset. After the SOM routine has clustered the multi-dimensional dataset, principal component regressions are derived between C_T , A_T and the SHPs using data within each neuron, each of which can be thought of as a local-scale optimizer that follows the global nonlinear optimization analysis performed by the SOM. To then predict C_T and A_T using any independent set of SHP measurements, a similarity measure is first used to determine which neuron best represents the SHP measurements, and then the C_T and A_T values are predicted using the regression parameters established using training data from that neuron. We call this approach SOMLO: self-organizing multiple linear output¹⁶. Once the monthly surface-ocean climatology for C_T and A_T was determined, p_{CO_2} was calculated using standard CO_2 dissociation constants.

SOMLO errors. The approach used here to determine p_{CO_2} distributions includes both systematic and random sources of error. The main source of random error is uncertainties within the global C_T and A_T distributions, which have been estimated to be $\pm 11.8 \mu\text{mol kg}^{-1}$ and $\pm 10.2 \mu\text{mol kg}^{-1}$, respectively¹⁶. Excluding the Arctic Ocean (north of 70°N), errors were estimated to be $\pm 10.9 \mu\text{mol kg}^{-1}$ and $\pm 9 \mu\text{mol kg}^{-1}$, respectively, for C_T and A_T . To quantify the corresponding uncertainty in calculated p_{CO_2} values, we used the independent C_T and A_T predictions of ref. 16 to calculate p_{CO_2} values, and then compared these values to those calculated using the *in situ* C_T and A_T measurements. This independent testing revealed global uncertainties in calculated p_{CO_2} of about 27 μatm ($N = 16,727$), or 23 μatm when excluding regions north of 70°N. We also compared our independent SOMLO p_{CO_2} predictions to the bottle-derived measurements at the 18-year Bermuda Atlantic time-series station and found SOMLO captured the structure and variability of the seasonal cycle well¹¹. Although there were twice as many global summertime measurements for carbon as there were for winter, we found no seasonal bias from our independent testing of SOMLO¹⁶.

Revelle factor and p_{CO_2} amplification. The Revelle factor (R) is defined by the relative change in p_{CO_2} induced by a change in total dissolved inorganic carbon: $R = \frac{\partial \ln(p_{CO_2})}{\partial \ln(C_T)}$. Revelle factors are lowest in the subtropical oceans ($R \approx 8$) and highest in the polar oceans ($R \approx 15$); see Extended Data Fig. 1. The uptake of anthropogenic CO_2 over decadal scales increases C_T while A_T remains constant. Meanwhile, diurnal to monthly variability also changes C_T . Future p_{CO_2} levels are nonlinearly amplified by a combination of long-term decadal-scale anthropogenic CO_2 increases to the Revelle factor and shorter-term seasonal-scale changes in C_T via natural processes (for example, biological processes or mixing dynamics).

SOMLO-based future projections. All future projections for p_{CO_2} from our SOMLO data-based analysis exclude those regions where the annual amplitude of the annual cycle of CO_2 is less than our known global uncertainty of 27 μatm . We use our data-derived SOMLO reconstructions of present-day C_T variability to calculate future p_{CO_2} patterns assuming a constant air-sea disequilibrium with rising atmospheric CO_2 . Although using our data-based patterns of C_T and A_T in the year 2000 provides a good constraint on observed variability, projections of future CO_2 using our SOMLO approach have to incorporate three steady-state assumptions: that the air-sea CO_2 disequilibrium is unchanged from the year 2000; that the

observed seasonal variability in both C_T and A_T does not change in the future; and that the effects of climate-warming feedbacks are negligible. On the basis of these assumptions, we expect the SOMLO predictions to be an upper-bound estimate for future p_{CO_2} ; however, we use a climate model to quantify the likely bias.

Quantifying biases in the SOMLO approach. We took C_T and A_T output from the HadGEM2-ES model¹⁴ for the year 2006. We then calculated future monthly p_{CO_2} concentrations until 2100 assuming the RCP8.5 scenario using the same steady-state assumptions used in our SOMLO analysis. This allowed us to compare the steady-state future p_{CO_2} calculations with the model output that incorporates air-sea gas exchange, mixing and climate effects, to quantify the likely bias when applying steady-state assumptions in our SOMLO approach. The results are shown in Extended Data Fig. 2, with a global bias of approximately 2% by 2100. However, in some regions in the North Atlantic, North Pacific and Southern oceans, we find that the SOMLO steady-state approach is likely to under-estimate future p_{CO_2} by up to 10%, whereas in other regions, such as the equatorial Pacific and high latitudes, the steady-state SOMLO approach overestimates p_{CO_2} by about 10%–15%.

Climate models. To explore the effect of increasing CO_2 air-sea disequilibrium and climate-change feedbacks (biologically, temperature- or mixing-driven) on future changes in the ocean carbon system, we used the output from four ESMs participating in the Coupled Model Inter-comparison Project (CMIP5; see Extended Data Table 1 for list of models). Each model was forced with changes in climate and atmospheric CO_2 (according to RCP8.5), and incorporated a dynamic ocean biogeochemical model with ocean C_T , A_T , temperature and salinity as state variables for output. For robust comparison, p_{CO_2} was calculated using the equivalent CO_2 dissociation constants as used for the data-based method. All models predict that specific regions will undergo substantial CO_2 amplification by 2100, ranging between two- and 12-fold p_{CO_2} amplification (Extended Data Fig. 3). However, there are considerable spatiotemporal variations between models that need to be explored in the future (Figs 2, 4 and Extended Data Fig. 3).

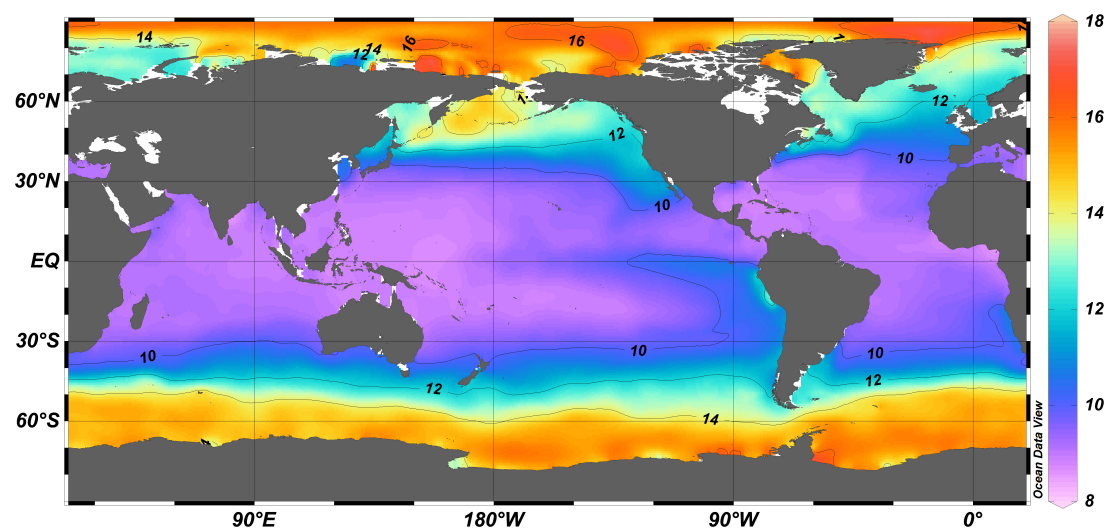
Model caveats. Global ocean models provide robust insights into how complex biogeochemistry, air-sea gas exchange and mixing are likely to alter future CO_2 projections. However, global climate models do not generally provide good constraints on smaller-scale and higher-frequency variability of CO_2 in the ocean. Recent studies suggest that global ocean models are considerably limited in their ability to constrain both the spatiotemporal patterns and magnitude of seasonal/natural CO_2 dynamics, particularly in the Southern, North Pacific, Indian and Northern (subpolar) Atlantic oceans^{26–29}. To investigate this on a global basis, we perform a statistical analysis to compare how ESMs perform in capturing the observed magnitude and dynamics of the annual CO_2 cycle (Extended Data Fig. 4). All models show low correlation (0.2–0.4) with global patterns of observed CO_2 variability and substantially under-estimate the observed magnitude of CO_2 variability. Annual observed variability of CO_2 is estimated to be about 65 μatm , whereas models predict a variability of about 38 μatm . This under-representation of contemporary p_{CO_2} variability in models will directly result in an underestimation of the extent of future month-long hypercapnia events by models.

Interannual variability. We explored the effect of interannual variability on future CO_2 amplification within one of the models (Extended Data Fig. 5). We find localized variations in the patterns and magnitude of future CO_2 amplification (depending on the model dynamics), however, the overall larger-scale patterns of up to tenfold future CO_2 amplification is robust on interannual timescales.

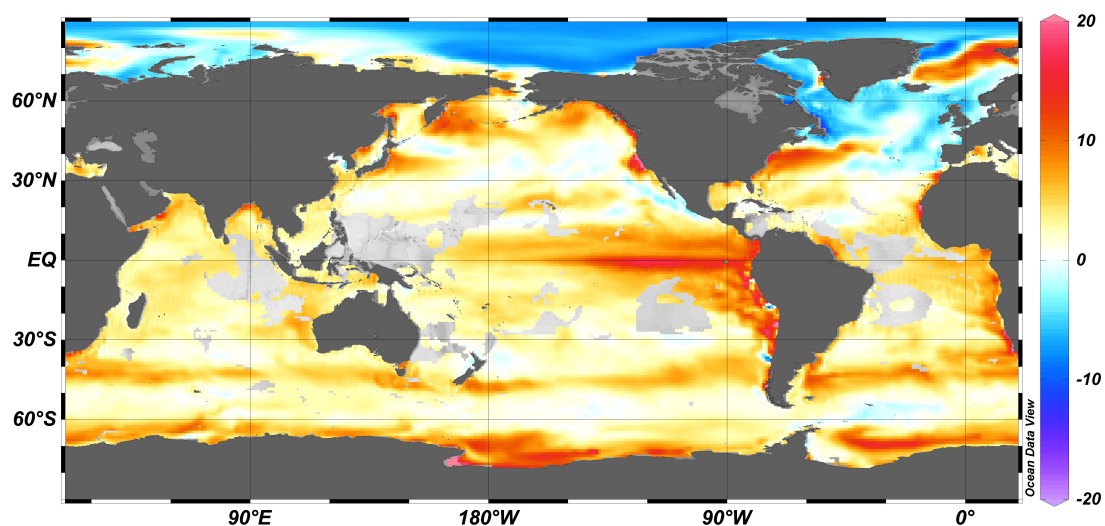
Robustness of nonlinearity of future p_{CO_2} amplification in models. A linear amplification of the annual p_{CO_2} cycle implies that the lowest monthly p_{CO_2} concentration will be reduced in an equal but opposite way to the highest monthly p_{CO_2} concentration in the ocean. However, what we show here with our SOMLO analysis is that the future amplification is nonlinear, with higher p_{CO_2} values leading to a greater amplification than lower values (Fig. 2). To investigate whether this nonlinearity is a feature of the global climate models, we plot normalized p_{CO_2} distributions for two climate models and compare them with SOMLO (Extended Data Fig. 6). We find that both distributions show nonlinear amplification (greater for higher p_{CO_2} values), suggesting this nonlinearity to be a robust feature of our findings.

32. Key, R. M. *et al.* A global ocean carbon climatology: results from global data analysis project (GLODAP). *Glob. Biogeochem. Cycles* **18**, GB4031 (2004).

33. Kohonen, T. *Self-organizing Maps* 3rd edn (Springer, 2001).

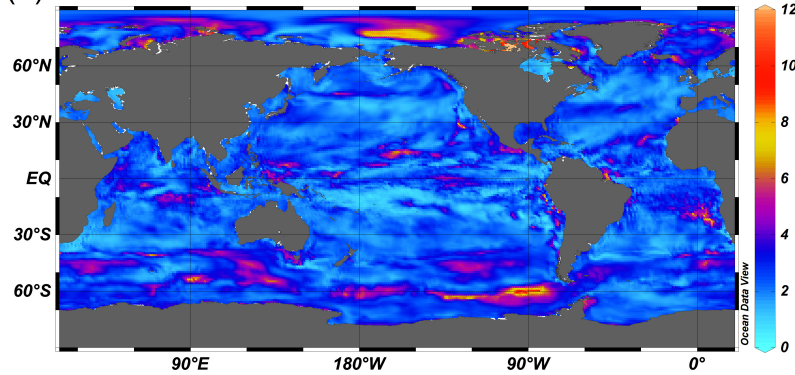


Extended Data Figure 1 | Ocean Revelle factor. Distribution of the mean Revelle or ‘buffer’ factor in the surface ocean in 2000, determined from our study.

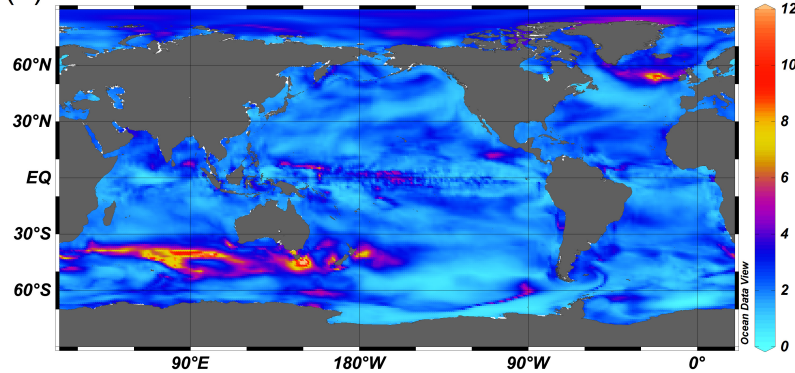


Extended Data Figure 2 | Quantifying biases in the SOMLO approach. Bias in p_{CO_2} concentrations (expressed as a percentage) in 2100 due to using steady-state assumptions in the SOMLO analysis. Positive (negative) values indicate that SOMLO over-estimates (under-estimates) p_{CO_2} .

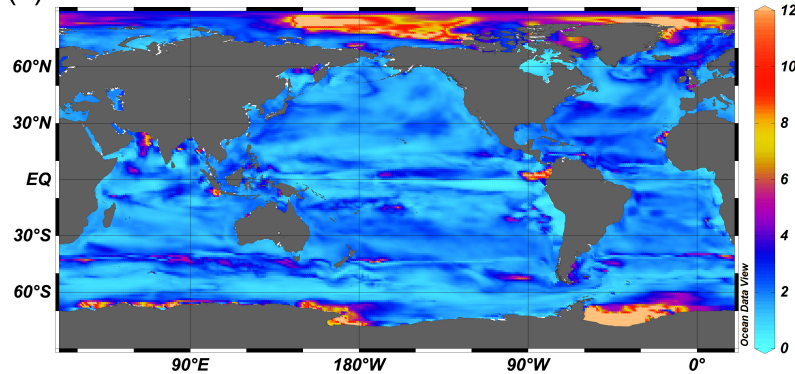
(a) HadGem2-ES



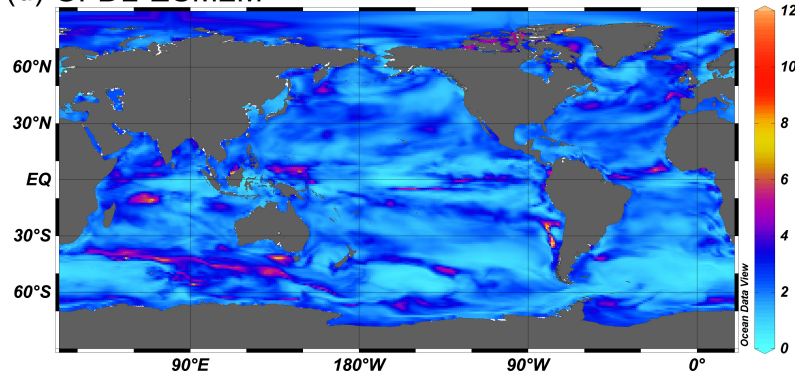
(b) IPSL-CM5A-MR



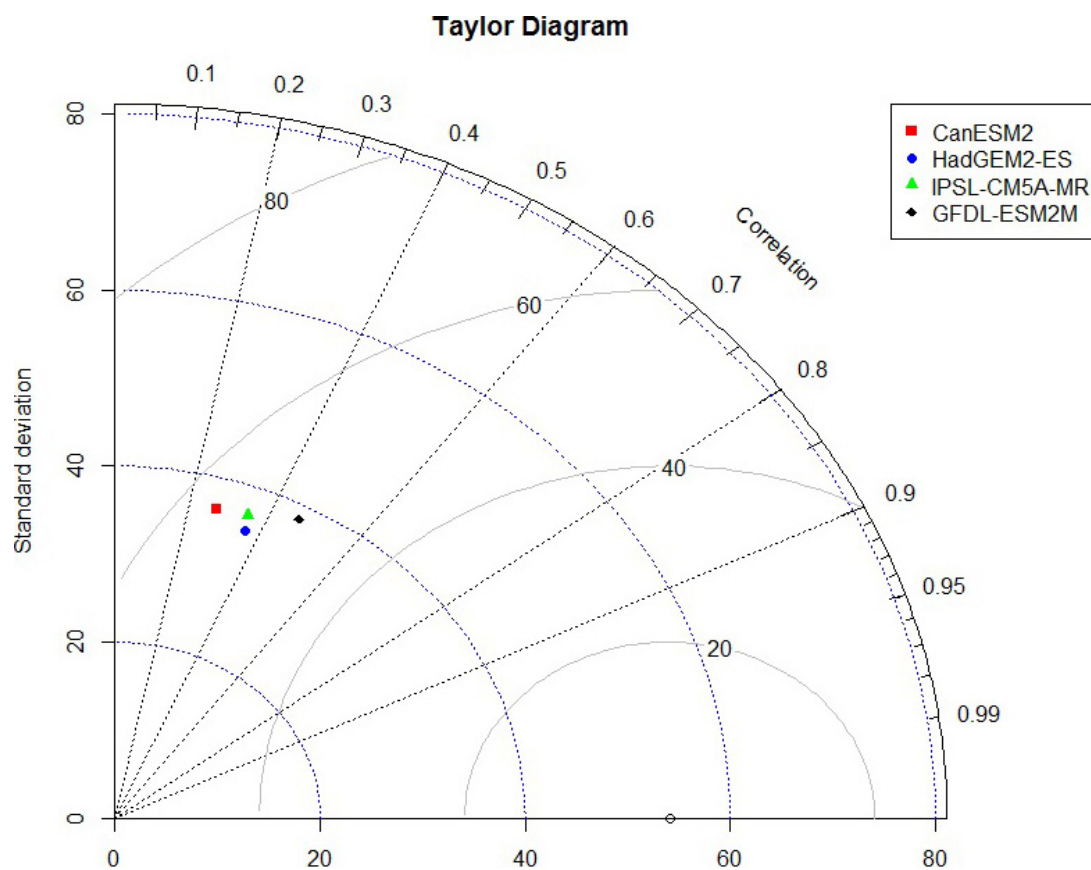
(c) CanESM2



(d) GFDL-ESM2M

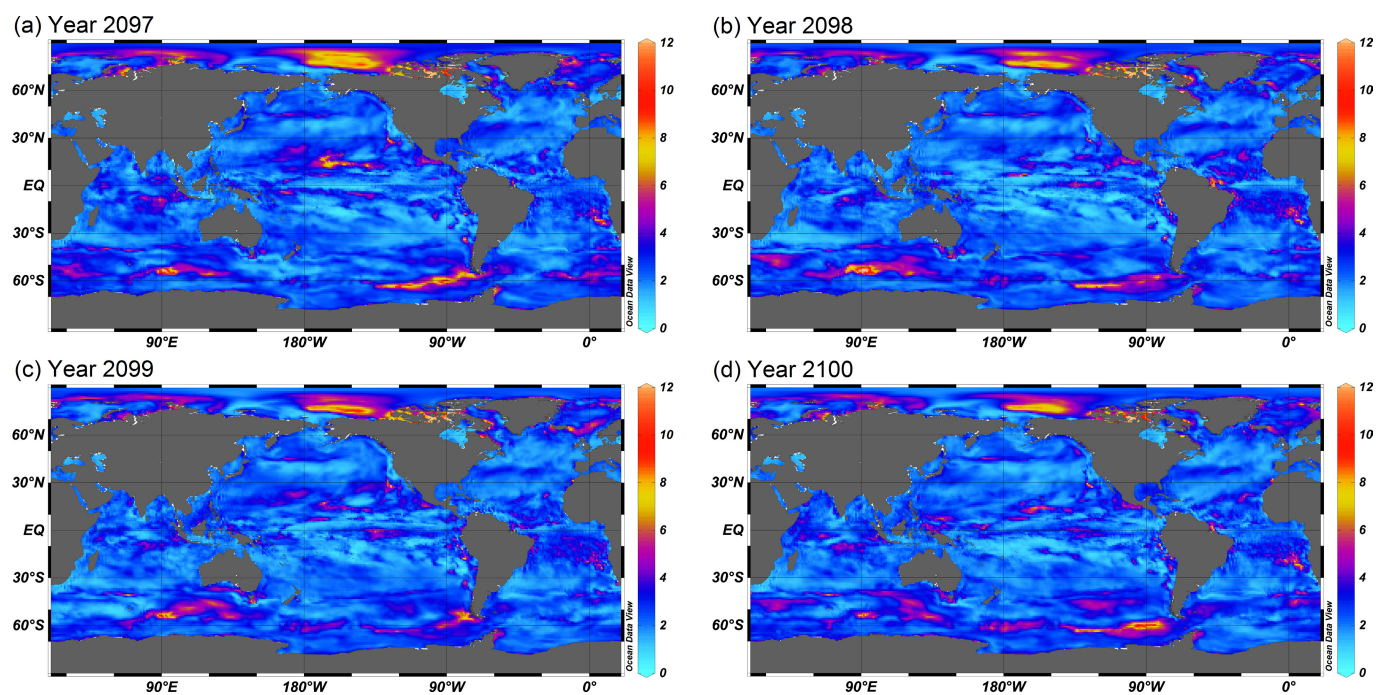


Extended Data Figure 3 | Future CO₂ amplification as predicted by climate models. a–d, CO₂ amplification factors by the year 2100 as predicted by four ESMs: HadGEM2-ES (a), IPSL-CM5A-MR (b), CanESM2 (c) and GFDL-ESM2M (d); see Extended Data Table 1 for model details.

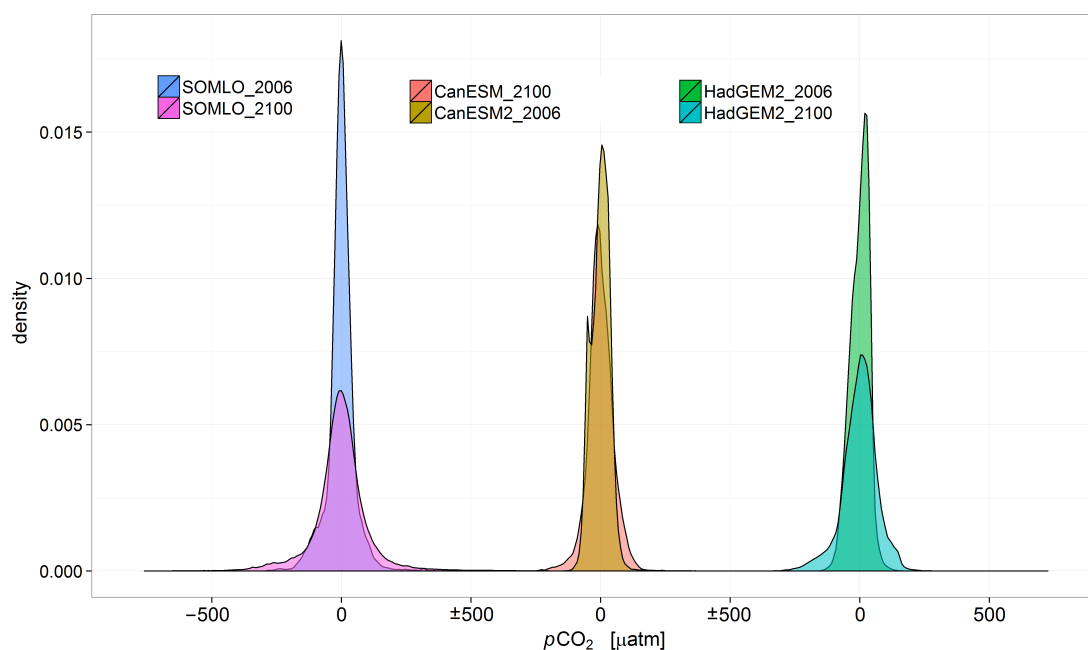


Extended Data Figure 4 | Comparison of present-day natural CO₂ variability as predicted by model and data. Taylor diagram comparing the ability of ESMs to capture the dynamics and magnitude of the annual cycle of CO₂ with our data-based approach. Here, the open circle on

the x axis represents the standard deviation in our data-based predictions, while the grey contour lines represent the residual standard error between the ESMs and our data-based predictions.



Extended Data Figure 5 | Interannual variability of future CO₂ amplification. CO₂ amplification factors within the HadGEM2 model for each year from 2097 to 2100.



Extended Data Figure 6 | Nonlinearity of future CO₂ distributions between SOMLO and models. Surface p_{CO_2} distribution (normalized such that the mean p_{CO_2} level across the whole surface ocean is shifted to zero) as predicted by SOMLO and two climate models (CanESM2 and

HadGEM2-ES), highlighting the change in distribution between 2006 and 2100. The nonlinearity is evidenced by ‘bulging’ in the 2100 distributions for positive p_{CO_2} , relative to those for negative p_{CO_2} .

Extended Data Table 1 | Main characteristics of the four ESMs used in this study

Model	Ocean Resolution	Biogeochemical model	Reference
CanESM2			15
HadGEM2-ES	0.3-1°	Diat-HadOCC	14
IPSL-CM5A-MR	0.5-1°	PISCES	12
GFDL-ESM2M	0.4°		13

Models are presented in refs 12–15.

Powering Earth's dynamo with magnesium precipitation from the core

Joseph G. O'Rourke¹ & David J. Stevenson¹

Earth's global magnetic field arises from vigorous convection within the liquid outer core. Palaeomagnetic evidence reveals that the geodynamo has operated for at least 3.4 billion years¹, which places constraints on Earth's formation and evolution. Available power sources in standard models include compositional convection (driven by the solidifying inner core's expulsion of light elements), thermal convection (from slow cooling), and perhaps heat from the decay of radioactive isotopes. However, recent first-principles calculations^{2,3} and diamond-anvil cell experiments^{4,5} indicate that the thermal conductivity of iron is two or three times larger than typically assumed in these models. This presents a problem: a large increase in the conductive heat flux along the adiabat (due to the higher conductivity of iron) implies that the inner core is young (less than one billion years old⁴), but thermal convection and radiogenic heating alone may not have been able to sustain the geodynamo during earlier epochs. Here we show that the precipitation of magnesium-bearing minerals from the core could have served as an alternative power source. Equilibration at high temperatures in the aftermath of giant impacts allows a small amount of magnesium (one or two weight per cent) to partition into the core while still producing the observed abundances of siderophile elements in the mantle and avoiding an excess of silicon and oxygen in the core. The transport of magnesium as oxide or silicate from the cooling core to underneath the mantle is an order of magnitude more efficient per unit mass as a source of buoyancy than inner-core growth. We therefore conclude that Earth's dynamo would survive throughout geologic time (from at least 3.4 billion years ago to the present) even if core radiogenic heating were minimal and core cooling were slow.

Earth differentiated into a silicate mantle and an iron-rich core during its formation in the approximately 100 million years following the initial collapse of the solar nebula. Scores of collisions between planetary embryos ranging from Moon- to Mars-sized accompanied by an influx of smaller planetesimals characterized the last stage of accretion^{6,7}. In one view of core formation, impacting embryos disintegrate in a deep magma ocean that overlies a mostly solid region^{8–12}. Metal sinks into a pond at the base of the magma ocean and equilibrates at the ambient pressure and temperature near the peridotite liquidus. As a consequence, the mantle becomes depleted in siderophile elements. Metallic diapirs quickly sink to the core–mantle boundary (CMB) without further equilibration, although percolation through the lower mantle has also been suggested¹³. One issue with this view is that the concentration of light elements in core material should increase over time as pressure and temperature increase. Without vigorous mixing, a stable compositional stratification would develop below the CMB¹⁴, which could prevent the initialization of a convective dynamo. The Moon-forming giant impact and possibly earlier, less energetic ones may recover the homogeneity of the core. Giant impacts are also important because they can heat parts of Earth to around 10,000 K, possibly even causing complete melting^{15–17}. Portions of the cores of large impactors should emulsify down to short (centimetres) length scales in

the superheated mantle¹⁸, permitting some metal–silicate equilibration at temperatures far above the peridotite liquidus.

The density of the core today, inferred from seismology, is about 10% less than the expected value for a pure iron/nickel (Fe/Ni) alloy at the high temperatures and pressures of Earth's deep interior¹⁹. Identifying the light element(s) responsible for this deficit has remained a central problem for decades. Despite being the fourth most abundant element in Earth by mass, magnesium (Mg) has been correctly rejected as a major contributor to the density deficit of the core in favour of elements like S, Si, O, C, and H, since Mg (as an oxide or silicate) and metallic Fe are almost immiscible under ambient conditions¹⁹.

However, the complete absence of Mg from the core is a thermodynamic impossibility because of entropic effects. New first-principles calculations indicate that the saturation limit of MgO in Fe is below 0.1 mol% at 3,000 K and 50 GPa, but reaches 1 mol% at 4,200 K and increases rapidly thereafter²⁰. Partitioning studies in diamond-anvil cells likewise indicate that Mg is not present above the detection limit after equilibration below 3,000 K. But Mg abundances approaching 1 mol% are observed in experiments reaching 3,000 K to 5,000 K (refs 21, 22). Spurious signals from the surrounding silicate are a complicating factor in these experiments, given the small sample sizes²². Still, all available data are consistent with the solubility of Mg in the Fe-rich metal exhibiting an exponential (Arrhenius) dependence on temperature. The low solubility also necessarily implies a strong dependence of that solubility on temperature.

Here we demonstrate that the composition of the mantle is consistent with core–mantle differentiation partially occurring at temperatures that permit substantial Mg partitioning. We consider two simple models of Earth's accretion, assuming that Earth is comprised of 12 major (Fe, Si, and O), lithophile (Mg, Al, and Ca), and siderophile (Ni, Co, Cr, V, Nb, and Ta) elements. First, we test a conventional, single-stage model with equilibration temperatures near the peridotite liquidus at mid-mantle depths. Second, we develop a two-stage model with some material equilibrating at temperatures above 5,000 K. To calculate the composition of Earth's core and mantle, we adopt parameterizations of partition coefficients (the metal-to-silicate ratio of elemental concentrations) as functions of pressure, temperature, and bulk composition from the most recent and comprehensive data sets^{12,22}. We compare the predictions of our models to the estimated composition of the primitive mantle and the core mass fraction²³.

A Markov chain Monte Carlo analysis quantifies the distribution of model parameters that match the available data. This technique explores the effects of varying all variables within their plausible uncertainties, with tens of orders of magnitude more computational efficiency than a simple grid search of the high-dimensional parameter space. The median chi-squared values are 8.3 and 8.6 for the single- and two-stage models, respectively, indicating that both models can provide good fits to the data. Figure 1 shows that our two-stage model reproduces the observed elemental abundances in the primitive mantle. That is, the peaks of the probability density functions are all within 0.83σ of the estimated mean values. The median result of this model is

¹Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA.

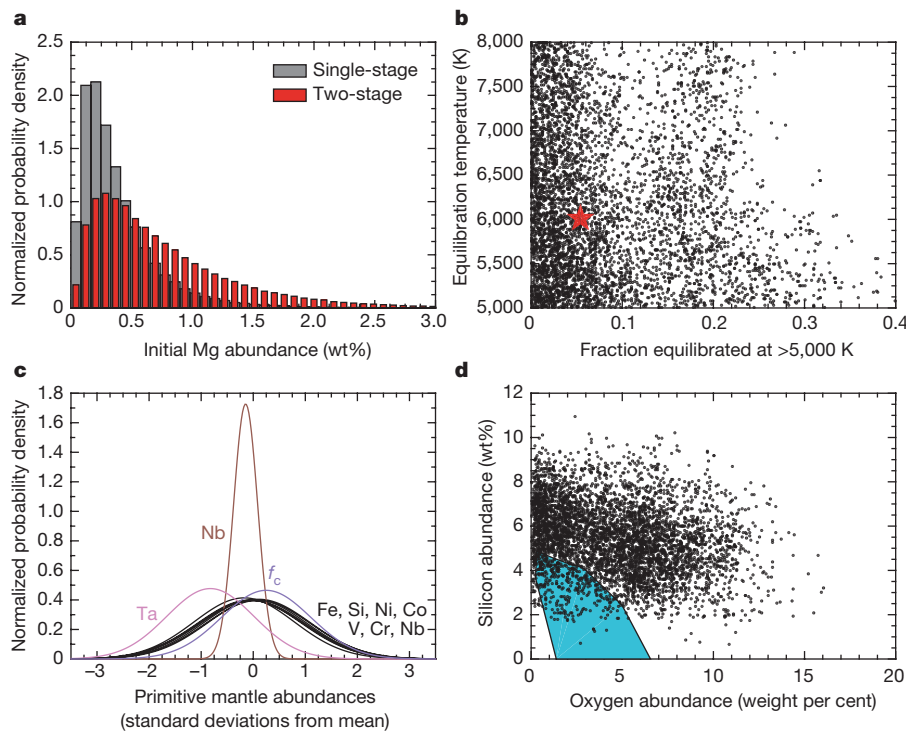


Figure 1 | Composition of Earth's core and mantle immediately after accretion from models of silicate-metal equilibrium. **a**, Initial abundance of Mg in the core from both models. **b–d**, From the two-stage model: fraction of material (red star is the mean result) equilibrated in the aftermath of giant impacts (**b**), posterior probability densities for elemental abundances in the primitive mantle²³ and the core mass fraction, f_c (**c**), and the initial abundances of Si and O in the core compared to those permitted (blue region) by present-day seismological observations²⁴ (**d**). Black dots are random draws from the posterior distributions.

that ~5 wt% of Earth equilibrates at ~6,000 K, although the equilibration of up to ~20 wt% at 8,000 K is permissible if other parameters are suitably adjusted. Mirroring previous studies, our single-stage model suggests a best-fit equilibration pressure of 57 ± 9 GPa, corresponding to $3,500 \pm 250$ K. Without partial equilibration at higher temperatures, less than 0.5 wt% Mg is expected to enter the core. The two-stage model, however, permits initial Mg abundances ranging from ~0.5 wt% to 2 wt%.

Other light elements also enter the core as temperatures of equilibration increase. Figure 1 illustrates that siderophile elements alone do not tightly constrain the initial abundances of Si and O in the core, especially since calculating partitioning behaviour above 5,000 K necessitates extrapolating the available experimental data²². Some two-stage simulations yield compositions within the 1σ range of estimated values for the present-day core based on seismology and mineral physics²⁴. But our models seem to 'prefer' slightly higher initial abundances of light elements. Precipitation caused by rapidly decreasing Mg solubility in the cooling core can lower the abundances of Si and, especially, O to their modern values.

Figure 2 shows representative calculations of the precipitation of Mg-bearing silicates and oxides from Earth's core. Here we assume that the core initially contained 3 wt% Si and 6 wt% O. We find that the amount of mass precipitated as a function of temperature is a sensitive function of the solubility of O, along with the expected dependence on the initial Mg abundance. Using the mean exchange coefficient for Mg, at least 1 wt% Mg is required for precipitation to begin above 4,000 K. Alternatively, the constants used to calculate the exchange coefficient for O could be decreased by more than 0.25σ from their estimated mean values. Extended Data Fig. 1 illustrates one example of the evolving composition of the precipitate. The solubility of Mg is below that of Si and O, so it comprises almost half the precipitate by number despite having the lowest abundance. The precipitate becomes more Si-rich as Mg is depleted, whereas the Fe content remains roughly constant. Overall, ~0.5% of the initial mass precipitates per 100 K of cooling once precipitation begins.

Transporting precipitated material to the CMB from a mostly well-mixed outer core releases more gravitational energy per unit mass than inner-core growth. Large-scale vertical motions are necessary to maintain the dynamo, meaning that either thermal or compositional

effects must provide enough buoyancy to keep the temperature profile in the liquid outer core close to the adiabat in the region of dynamo generation^{25,26}. Crucially, the density difference between the precipitate and the outer core is approximately ten times the contrast between the inner and outer core attributable to composition (roughly half of the ~5% total). This means that precipitating a layer of Mg-bearing material with a thickness of only ~10 km above the CMB is energetically equivalent to crystallizing the entire inner core. We incorporate Mg precipitation into comprehensive models of the energetics of the core to explore its importance further.

Figure 3 presents several calculations for the thermochemical evolution of the core. We iterate the equations of global energy and entropy balance backwards in time from present-day conditions²⁶. Given a fixed amount of entropy production, we calculate the CMB heat flow and cooling rate necessary to sustain the dynamo. We use the recently

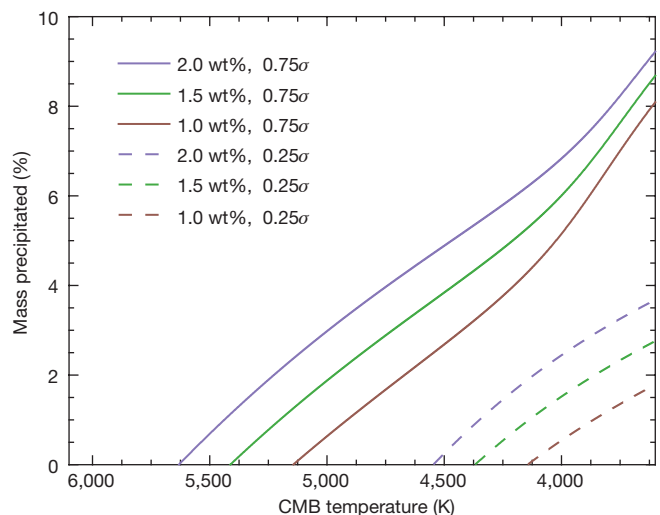


Figure 2 | Estimates of mass precipitated from the cooling core. The colours denote initial Mg abundances, while solid and dashed lines represent calculations with the constants a_O , b_O , and c_O used to calculate O solubility (K_O) decreased by different fractions of a standard deviation.

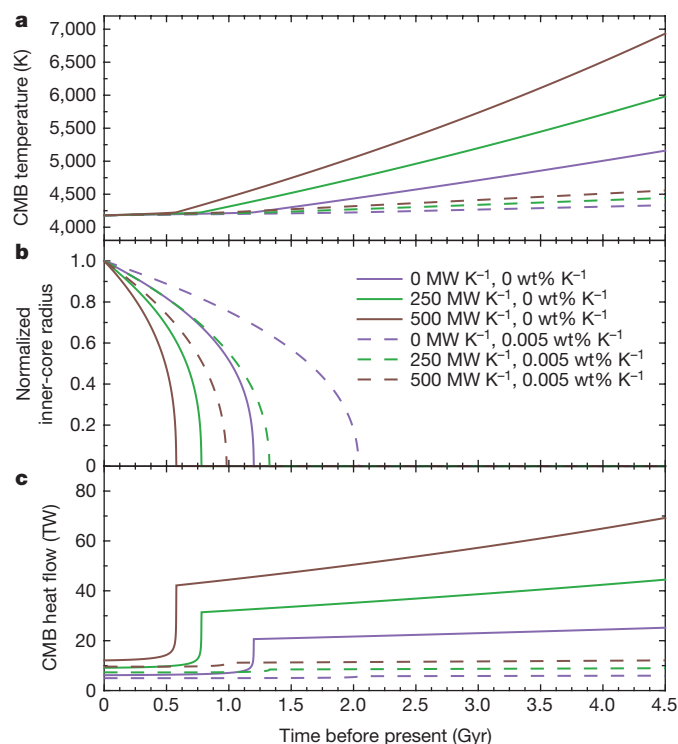


Figure 3 | Thermochemical evolution of the core for various rates of precipitation and entropy production associated with ohmic dissipation. Assuming that the core always produces a constant amount of entropy required to sustain the dynamo, we calculate the implied CMB temperature (a), inner-core radius relative to the present (b), and CMB heat flow (c). Gyr, billion years.

revised value of core thermal conductivity, despite a new study²⁷ that militates against the emerging theoretical and experimental consensus^{2–5}. If there is no Mg precipitation, an entropy production rate of 500 MW K⁻¹ (corresponding to ~2.5 TW of ohmic dissipation) implies initial CMB temperatures of ~7,000 K and a core-to-mantle heat flow of 40 TW to 70 TW before inner-core nucleation at ~0.6 billion years ago. Incorporating the plausible precipitation rate lowers the necessary amount of secular cooling to only ~300 K over 4.5 billion years. A core heat flow close to the present-day value (<20 TW) is thus sufficient to power a dynamo for Earth's entire history.

Precipitation of Mg from the core has profound implications for the evolution of Earth's deep interior. Most importantly, it eliminates the need to invoke a geochemically dubious magnitude of radiogenic heating²⁸ or enhanced heat flux across the CMB into a basal magma ocean²⁹. High thermal conductivity and slow core cooling are consistent with inner-core nucleation in the Mesoproterozoic³⁰. Models that include only the inner core as a source of compositional buoyancy predict that stable layers hundreds of kilometres thick should develop near the CMB³, which may be disrupted by precipitation. However, precipitation may actually occur at depth if the solubility of Mg is strongly pressure dependent. The real situation is even more complicated if the CMB is undersaturated in Si and O, meaning that material from the mantle tends to dissolve in the core. Elemental transport in both directions is potentially permissible because the Mg-rich precipitate differs in composition from the CMB. The effect of giant impacts on core formation should motivate additional experiments on metal–silicate partitioning at temperatures above 5,000 K. Non-standard evolutionary scenarios featuring precipitation are perhaps applicable to the cores of other terrestrial planets.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 January; accepted 26 November 2015.

1. Tarduno, J. A. *et al.* Geodynamo, solar wind, and magnetopause 3.4 to 3.45 billion years ago. *Science* **327**, 1238–1240 (2010).
2. de Koker, N., Steinle-Neumann, G. & Vlcek, V. Electrical resistivity and thermal conductivity of liquid Fe alloys at high *P* and *T*, and heat flux in Earth's core. *Proc. Natl Acad. Sci. USA* **109**, 4070–4073 (2012).
3. Pozzo, M., Davies, C., Gubbins, D. & Alfe, D. Thermal and electrical conductivity of iron at Earth's core conditions. *Nature* **485**, 355–358 (2012).
4. Gomi, H. *et al.* The high conductivity of iron and the thermal evolution of the Earth's core. *Phys. Earth Planet. Inter.* **224**, 88–103 (2013).
5. Seagle, C. T., Cottrell, E., Fei, Y., Hummer, D. R. & Prakapenka, V. B. Electrical and thermal transport properties of iron and iron-silicon alloy at high pressure. *Geophys. Res. Lett.* **40**, 5377–5381 (2013).
6. Chambers, J. E. Planetary accretion in the inner Solar System. *Earth Planet. Sci. Lett.* **223**, 241–252 (2004).
7. Oghara, M., Ida, S. & Morbidelli, A. Accretion of terrestrial planets from oligarchs in a turbulent disk. *Icarus* **188**, 522–534 (2007).
8. Wade, J. & Wood, B. J. Core formation and the oxidation state of the Earth. *Earth Planet. Sci. Lett.* **236**, 78–95 (2005).
9. Wood, B. J., Walter, M. J. & Wade, J. Accretion of the Earth and segregation of its core. *Nature* **441**, 825–833 (2006).
10. Rubie, D. C. *et al.* Heterogeneous accretion, composition and core-mantle differentiation of the Earth. *Earth Planet. Sci. Lett.* **301**, 31–42 (2011).
11. Siebert, J., Badro, J., Antonangeli, D. & Ryerson, F. J. Terrestrial accretion under oxidizing conditions. *Science* **339**, 1194–1197 (2013).
12. Rubie, D. C. *et al.* Accretion and differentiation of the terrestrial planets with implications for the compositions of early-formed Solar System bodies and accretion of water. *Icarus* **248**, 89–108 (2015).
13. Shi, C. Y. *et al.* Formation of an interconnected network of iron melt at Earth's lower mantle conditions. *Nature Geosci.* **6**, 971–975 (2013).
14. Helffrich, G. Outer core compositional layering and constraints on core liquid transport properties. *Earth Planet. Sci. Lett.* **391**, 256–262 (2014).
15. Canup, R. M. Accretion of the Earth. *Phil. Trans. R. Soc. A* **366**, 4061–4075 (2008).
16. Canup, R. M. Forming a Moon with an Earth-like composition via a giant impact. *Science* **338**, 1052–1055 (2012).
17. Cuk, M. & Stewart, S. T. Making the Moon from a fast-spinning Earth: a giant impact followed by resonant despinning. *Science* **338**, 1047–1052 (2012).
18. Dahl, T. & Stevenson, D. J. Turbulent mixing of metal and silicate during planet accretion—and interpretation of the Hf–W chronometer. *Earth Planet. Sci. Lett.* **295**, 177–186 (2010).
19. Poirier, J. Light elements in the Earth's outer core: a critical review. *Phys. Earth Planet. Inter.* **85**, 319–337 (1994).
20. Wahl, S. M. & Militzer, B. High-temperature miscibility of iron and rock in terrestrial planet formation. *Earth Planet. Sci. Lett.* **410**, 25–33 (2015).
21. Takafuji, N., Hirose, K., Mitome, M. & Bando, Y. Solubilities of O and Si in liquid iron in equilibrium with (Mg,Fe)SiO₃ perovskite and the light elements in the core. *Geophys. Res. Lett.* **32**, L06313 (2005).
22. Fischer, R. A. *et al.* High pressure metal–silicate partitioning of Ni, Co, V, Cr, Si, and O. *Geochim. Cosmochim. Acta* **167**, 177–194 (2015).
23. Palme, H. & O'Neill, H. in *Treatise on Geochemistry* 2nd edn (eds Holland, H. & Turekian, K.) 1–39 (Elsevier, 2013).
24. Badro, J., Cote, A. S. & Brodholt, J. P. A seismologically consistent compositional model of Earth's core. *Proc. Natl Acad. Sci. USA* **111**, 7542–7545 (2014).
25. Stevenson, D. J. Planetary magnetic fields. *Earth Planet. Sci. Lett.* **208**, 1–11 (2003).
26. Nimmo, F. in *Treatise on Geophysics* 2nd edn (ed. Schubert, G.) 31–65 (Elsevier, 2015).
27. Zhang, P., Cohen, R. E. & Haule, K. Effects of electron correlations on transport properties of iron at Earth's core conditions. *Nature* **517**, 605–607 (2015).
28. Corgne, A., Shantanu, K., Fei, Y. & McDonough, W. F. How much potassium is in the Earth's core? New insights from partitioning experiments. *Earth Planet. Sci. Lett.* **256**, 567–576 (2007).
29. Labrosse, S., Hernlund, J. W. & Coltice, N. A crystallizing dense magma ocean at the base of the Earth's mantle. *Nature* **450**, 866–869 (2007).
30. Biggin, A. J. *et al.* Palaeomagnetic field intensity variations suggest Mesoproterozoic inner-core nucleation. *Nature* **526**, 245–248 (2015).

Acknowledgements This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under grant number DGE-1144469 (J.G.O'R.).

Author Contributions J.G.O'R. performed the calculations and wrote the manuscript. D.J.S. designed the project, discussed the results, and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.G.O'R. (jorourke@caltech.edu).

METHODS

Element partitioning. For an element M besides oxygen with valence n , the partition coefficient is defined as $D_M = X_M^{\text{met}}/X_{\text{MO}_{n/2}}^{\text{sil}}$, where X_M^{met} is the mole fraction of the element M in metal and $X_{\text{MO}_{n/2}}^{\text{sil}}$ is the mole fraction of the corresponding oxide $\text{MO}_{n/2}$ in the silicate²². The associated exchange coefficient is $K_M = D_M/D_{\text{Fe}}^{n/2}$. For oxygen, $K_O = X_{\text{Fe}}^{\text{met}}X_{\text{O}}^{\text{sil}}/X_{\text{FeO}}$. Using experimental data, the exchange coefficients K_M may be parameterized as follows²²:

$$\begin{aligned} \log_{10} K_M = & a_M + \frac{b_M}{T} + \frac{c_M P}{T} + \frac{\varepsilon_M^{\text{M}} \ln(1 - X_M)}{2.303} \\ & + \frac{1}{2.303} \sum_{k=1}^N \varepsilon_k^i X_k \left(1 + \frac{\ln(1 - X_k)}{X_k} - \frac{1}{1 - X_i} \right) \\ & - \frac{1}{2.303} \sum_{k=1}^N \varepsilon_k^i X_k^2 X_i \left(\frac{1}{1 - X_i} + \frac{1}{1 - X_k} + \frac{X_i}{2(1 - X_i)^2} - 1 \right) \end{aligned}$$

where P is pressure, and a_M , b_M and c_M are constants. Each ε_k^i is the interaction parameter of elements i and j in the liquid at temperature T , which is a function of the reference interaction parameter e_j^i determined at a reference temperature of 1,873 K. These compositional parameters are necessary to fit partitioning data for V and Cr, but are typically set to zero for other siderophile elements²².

We estimate a_{Mg} and b_{Mg} using experimental results from ref. 21 and assuming that 0.05 wt% of Mg (below the detection limit) was present in quenched liquid iron at 2,500 K. Our derived values are in agreement with theoretical predictions²⁰ and preliminary results from diamond-anvil cell experiments at higher temperatures³¹. We adopt formal errors (100% and 9%, respectively) for these parameters a_{Mg} and b_{Mg} , which are comparable in magnitude to those for other elements²². We assume here that $c_{\text{Mg}} = 0$ and that interaction effects are negligible, although future experiments may reveal that these factors actually are important.

Equations for the exchange coefficients and mass balance allow us to solve for the composition of the metal and silicate phases for any initial bulk composition, given the temperature and pressure of equilibration. Methods for solving these equations are detailed by Rubie *et al.*^{10,12} and others^{11,22}. Since the valences of elements in the silicates are specified, oxygen fugacity is not a free parameter. Extended Data Table 1 lists experimentally determined values for all constants in our models, along with their formal errors.

Core formation. We consider two simple scenarios for Earth's accretion and differentiation. Both models have 37 parameters with quantified uncertainties corresponding to the oxygen-free composition of bulk Earth³² and the constants used to calculate partitioning behaviour^{12,22}. Our single-stage model has two additional parameters: P_1 , the pressure at equilibration, and $[\text{Fe}/\text{O}]_1$, the ratio of the molar abundances of iron and oxygen in bulk Earth. In this case, we assume that equilibration occurs at the base of a magma ocean with temperature on the peridotite liquidus. In our two-stage model, we permit a fraction of Earth to equilibrate at T_2 between 5,000 K and 8,000 K, which is hotter than the peridotite liquidus at the CMB. This material may equilibrate at a different pressure, P_2 , and can also have a different bulk oxygen content, $[\text{Fe}/\text{O}]_2$. We restrict P_1 and P_2 between 0 GPa and 135 GPa with equal prior probability throughout this range. Likewise, we allow $0.25 < [\text{Fe}/\text{O}]_1 < 0.65$ and $0.25 < [\text{Fe}/\text{O}]_2 < 0.6$, representing a range of oxidizing and reducing conditions. For reference, the best-fit $[\text{Fe}/\text{O}]_1 = 0.52 \pm 0.02$ in the single-stage model. We aim to reproduce ten data points: the estimated abundances of Fe, Si, Ni, Co, V, Cr, Nb and Ta in the primitive mantle, along with the ratio Nb/Ta (14.0 ± 0.3) and the core mass fraction, f_c (0.32 ± 0.01).

Given the high dimensionality of our models, a comprehensive grid search is computationally prohibitive. The Markov chain Monte Carlo technique, however, samples only each point in the parameter space that has a frequency equal to the posterior probability at that point. We use the standard Metropolis–Hastings sampling³³. Briefly, we begin with an initial guess for the model parameters, N_x . To decide whether to add another set of model parameters, N_{x+1} , to the chain, we compute the likelihood and prior probability. The posterior probability, $p(N_x)$, is proportional to the multiple of the likelihood and the prior. The likelihood is proportional to the conventional chi-squared value computed from the difference between the model output and the data. The prior is determined from experimental constraints. For example, a model that sets all constants used to calculate partitioning behaviour equal to their mean estimated values has a higher prior than a model where a few constants are increased or decreased by a fraction of a standard deviation. However, the second model may have a substantially higher likelihood and thus posterior. The algorithm always accepts new sets if $p(N_{x+1}) \geq p(N_x)$. Critically, new sets with lower posteriors are sometimes accepted with probability $p(N_{x+1})/p(N_x)$.

The generation functions that we used to guess N_{x+1} are tuned to produce an overall acceptance rate of ~25% to 75%. We report results from Markov chains with 1.66×10^6 and 2.71×10^6 accepted links for the single- and two-stage models, respectively. Visual inspection of the output traces reveals that these chains have converged, meaning that each Markov chain is equivalent to a long series of random draws from the real posterior distributions. To remove any dependence on our initial guess of model parameters, we remove all links before the likelihood first dips below its eventual median. Only 434 and 445 links were deleted from the Markov chains for the single- and two-stage models, respectively. Extended Data Fig. 2 shows the distributions of chi-squared values for both models, along with the posterior distributions for Earth's bulk composition and constants governing partitioning behaviour. As desired, posteriors for these parameters are basically equal to the priors. The posterior for the bulk abundance of Ta peaks 1.58 standard deviations away from the estimated mean, but values within 1σ remain quite plausible. Extended Data Fig. 3 contains the posterior probability distributions for other parameters in the two-stage model.

Magnesium precipitation. We constructed a simple model for the precipitation of Mg-rich silicates and oxides. The output of this calculation is the mass of precipitate associated with a given amount of cooling, which is an essential parameter in the complete thermochemical model detailed below. At a given temperature, a condition for silicates and oxides to precipitate may be written as:

$$X_{\text{FeO}} + X_{\text{MgO}} + X_{\text{SiO}_2} \geq 1$$

where X_i are the mole fractions of three metal oxides that are calculated as functions of the mole fractions and exchange coefficients for Fe, Mg, and O in the core as follows^{10–12,22}:

$$X_{\text{FeO}} = X_{\text{Fe}} \times X_{\text{O}} / K_{\text{O}}$$

$$X_{\text{MgO}} = X_{\text{FeO}} / X_{\text{Fe}} \times X_{\text{Mg}} / K_{\text{Mg}}$$

$$X_{\text{SiO}_2} = (X_{\text{FeO}} / X_{\text{Fe}})^2 \times X_{\text{Si}} / K_{\text{Si}}$$

Here we assume that precipitated material is instantaneously removed from the core. To generate Fig. 2, we begin with an undersaturated core at 6,000 K. We then decrease the temperature in increments of 1 K until precipitation begins. Next, we calculate how much of each oxide to remove to return the core to equilibrium. As temperature decreases further, the core gradually becomes depleted in light elements.

Thermochemical histories. Neglecting some small terms, the energy balance of the core is:

$$Q_{\text{CMB}} = Q_{\text{R}} + Q_{\text{S}} + Q_{\text{G}} + Q_{\text{G}'} + Q_{\text{L}}$$

where Q_{CMB} is the heat flux across the CMB, Q_{R} is the amount of radiogenic heating, Q_{S} is the magnitude of secular cooling, Q_{G} and $Q_{\text{G}'}$ are gravitational energy changes associated with inner-core growth and Mg precipitation, respectively, and Q_{L} is the latent heat released from phase transitions²⁶. Heating due to dynamo activity is both generated and dissipated within the core and thus is not included. It is, however, important to the entropy balance²⁶:

$$E_{\phi} + E_{\text{K}} = E_{\text{R}} + E_{\text{S}} + E_{\text{G}} + E_{\text{G}'} + E_{\text{L}}$$

Here E_{ϕ} , E_{K} and E_{R} are the entropy fluxes associated with ohmic dissipation, adiabatic heat flow, and radioactivity, respectively. The last four terms— E_{S} , E_{G} , $E_{\text{G}'}$ and E_{L} —are directly proportional to the rate of core cooling. We use parameterizations of these terms from Nimmo²⁶, plus our own estimates of the contribution from the precipitation of Mg-rich material detailed below. Other analytical models for the evolution of the core may imply higher initial CMB temperatures for dynamos driven by thermal convection alone³⁴. Throughout, we assume that the thermal conductivity equals $130 \text{ W m}^{-1} \text{ K}^{-1}$, based on the vast majority of recent studies^{2–5}, and that the abundance of potassium (K) in the core is 100 parts per million.

Combining the entropy and energy balances yields an expression for the heat flow across the CMB that is needed to drive a dynamo²⁶:

$$Q_{\text{CMB}} = Q_{\text{R}} \left[1 - \frac{Q_{\text{T}}}{E_{\text{T}}} \left(\frac{E_{\text{R}}}{Q_{\text{R}}} \right) \right] + \frac{Q_{\text{T}}}{E_{\text{T}}} (E_{\phi} + E_{\text{K}})$$

where the heat flow $Q_{\text{T}} = (Q_{\text{S}} + Q_{\text{G}} + Q_{\text{G}'} + Q_{\text{L}})/(dT_{\text{CMB}}/dt)$ and the entropy $E_{\text{T}} = (E_{\text{S}} + E_{\text{G}} + E_{\text{G}'} + E_{\text{L}})/(dT_{\text{CMB}}/dt)$ combine production terms that depend on the cooling rate. The temperature at the CMB is T_{CMB} . The implied cooling rate is $dT_{\text{CMB}}/dt = -(Q_{\text{CMB}} - Q_{\text{R}})/Q_{\text{T}}$. If we assume a fixed value for E_{ϕ} , we can iterate these equations backwards in time from the present to calculate the thermochemical history of the core. Increasing the amount of entropy required for the dynamo or the adiabatic heat flow implies higher initial temperatures and heat flow.

Radiogenic heating limits the amount of secular cooling, but actually mandates higher core-to-mantle heat flow to sustain a dynamo. In reality, the thermal evolution of the mantle determines Q_{CMB} and thus the amount of entropy available for the dynamo. By assuming that E_ϕ is fixed to a particular value, we are essentially computing the minimum heat flow that the mantle must be able to accommodate from the core.

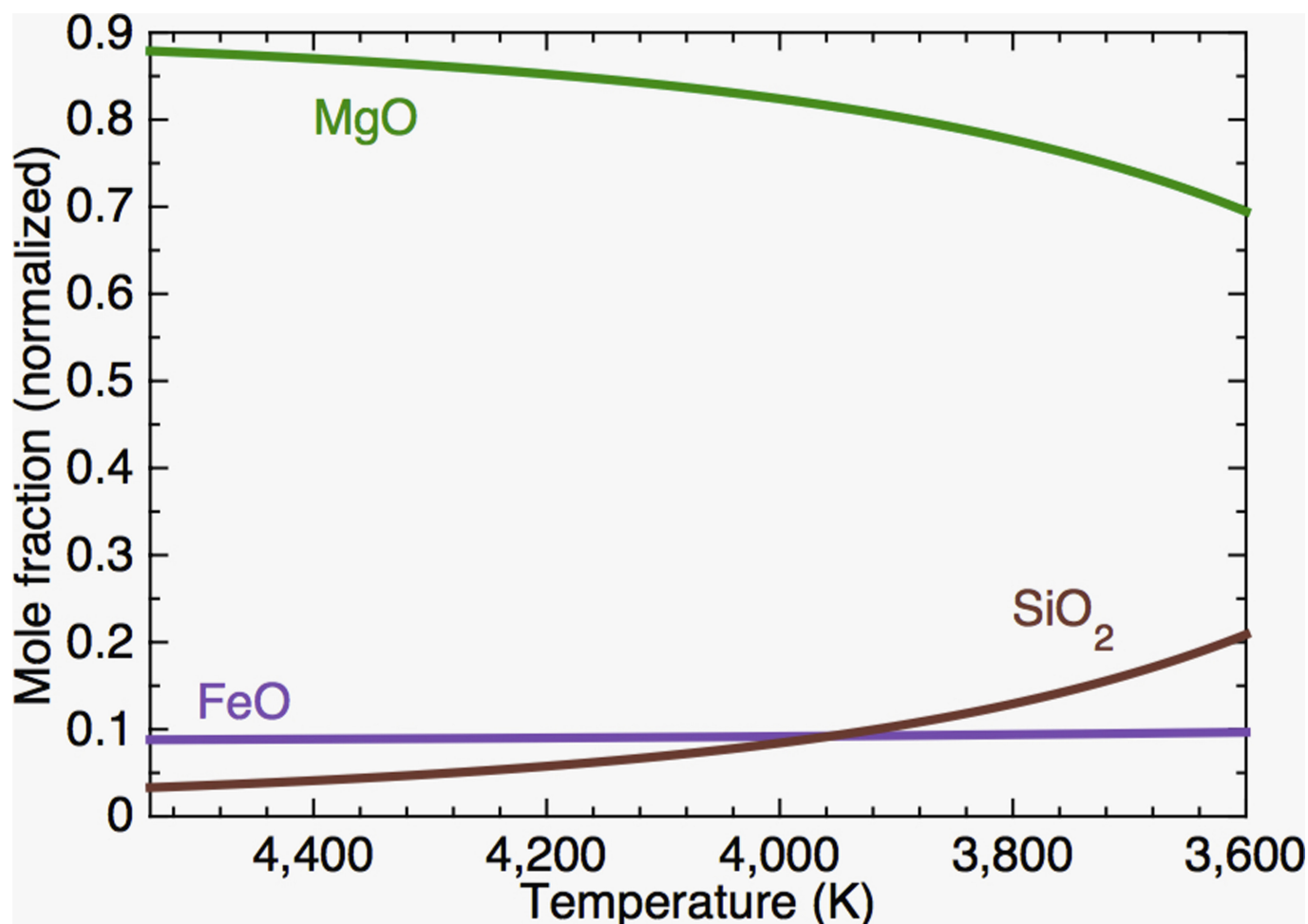
The gravitational contribution to core heating from Mg precipitation may be estimated as:

$$Q_{G'} = \int_{\infty} \varphi \left(\frac{\partial \rho}{\partial t} \right)_{P,T} dV = \int_{\infty} \varphi \rho \alpha_c \left[C_m \left(\frac{dT_{\text{CMB}}}{dt} \right) \right] dV$$

where φ is the gravitational potential, ρ is density, t is time, and V is volume. The coefficient of compositional expansivity $\alpha_c = -1/\rho(\partial \rho / \partial c)_{P,T} \approx 1.12$, where c is the concentration of the light elements that precipitate out of the core. The rate of precipitation, C_m , is normalized to the initial core mass and directly proportional to the rate of cooling dT_{CMB}/dt . Since inspection of Fig. 2 suggests that about 0.5% of initial core mass precipitates per 100 K of cooling, we set $C_m = 5 \times 10^{-5}$. The exclusion of Mg from the inner core, if any, is already taken into account by using

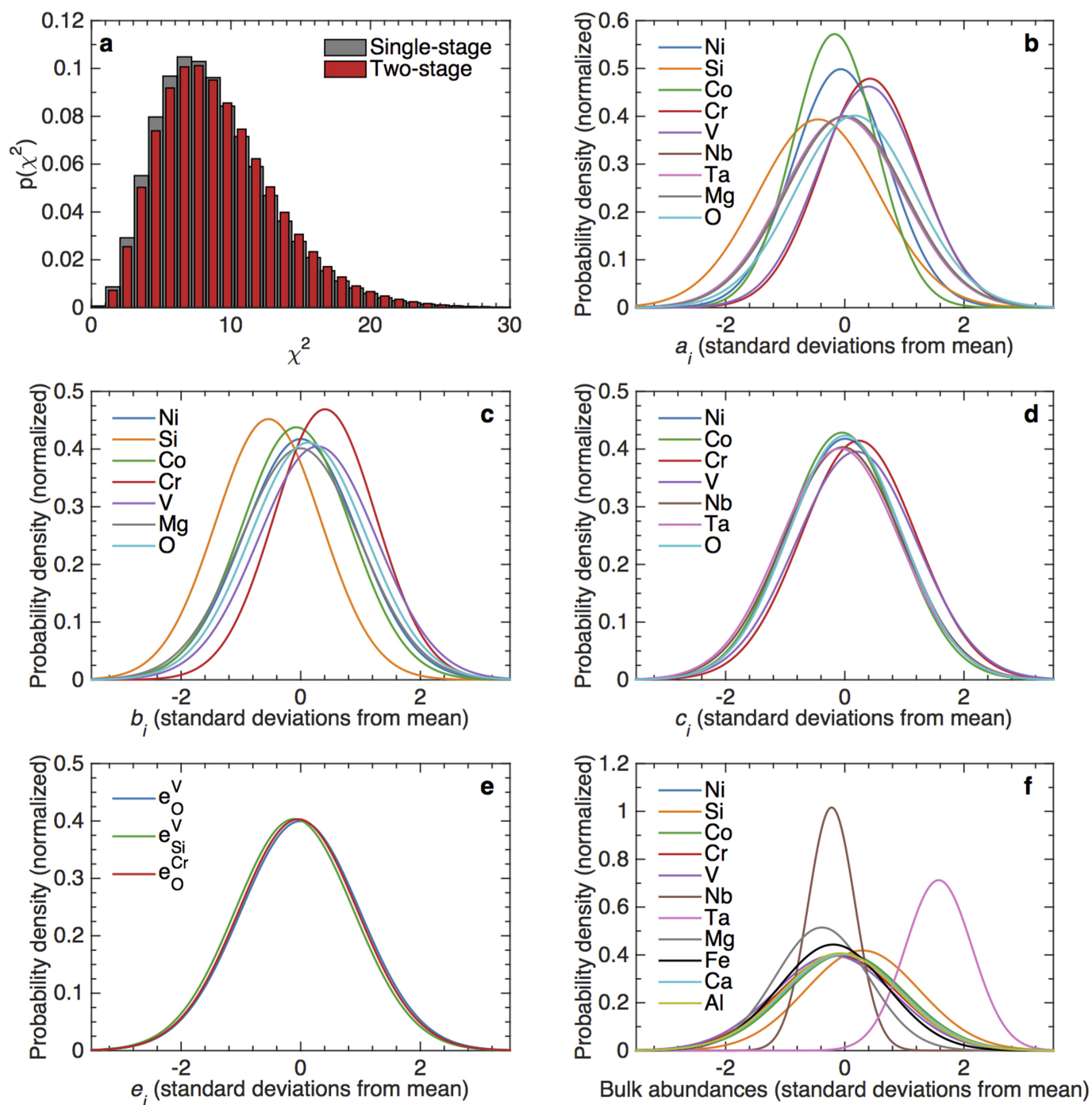
the empirical density difference across the inner-core boundary to calculate the gravitational contribution from the solidification of the inner core. The associated entropy flux²⁶ is simply $E_{G'} = Q_{G'}/T_{\text{CMB}}$. The entropy contribution from the latent heat of precipitation is exactly zero if precipitation occurs at the top of the core. Accordingly, we do not include any contribution from latent heat of precipitation in our calculation of Q_{CMB} . At present, $Q_L \approx 5$ TW and the rate of inner-core growth by mass is perhaps 2 to 4 times the amount of precipitation²⁶. Including the latent heat of precipitation would possibly increase Q_{CMB} by 1–2 TW and slightly decrease the implied amount of secular cooling.

31. Nomura, R. *et al.* Partitioning of potassium into the Earth's core and implications for thermal history of the Earth. *AGU Fall Meet. Abstr.* D133A–2411 (2012).
32. McDonough, W. F. in *Earthquake Thermodynamics and Phase Transformation in the Earth's Interior* (eds Teisseyre, R. & Majewski, E.) 3–23 (Academic Press, 2001).
33. Chib, S. & Greenberg, E. Understanding the Metropolis-Hastings algorithm. *J. Am. Stat. Assoc.* **49**, 327–335 (1995).
34. Labrosse, S. Thermal evolution of the core with a high thermal conductivity. *Phys. Earth Planet. Inter.* **247**, 36–55 (2015).



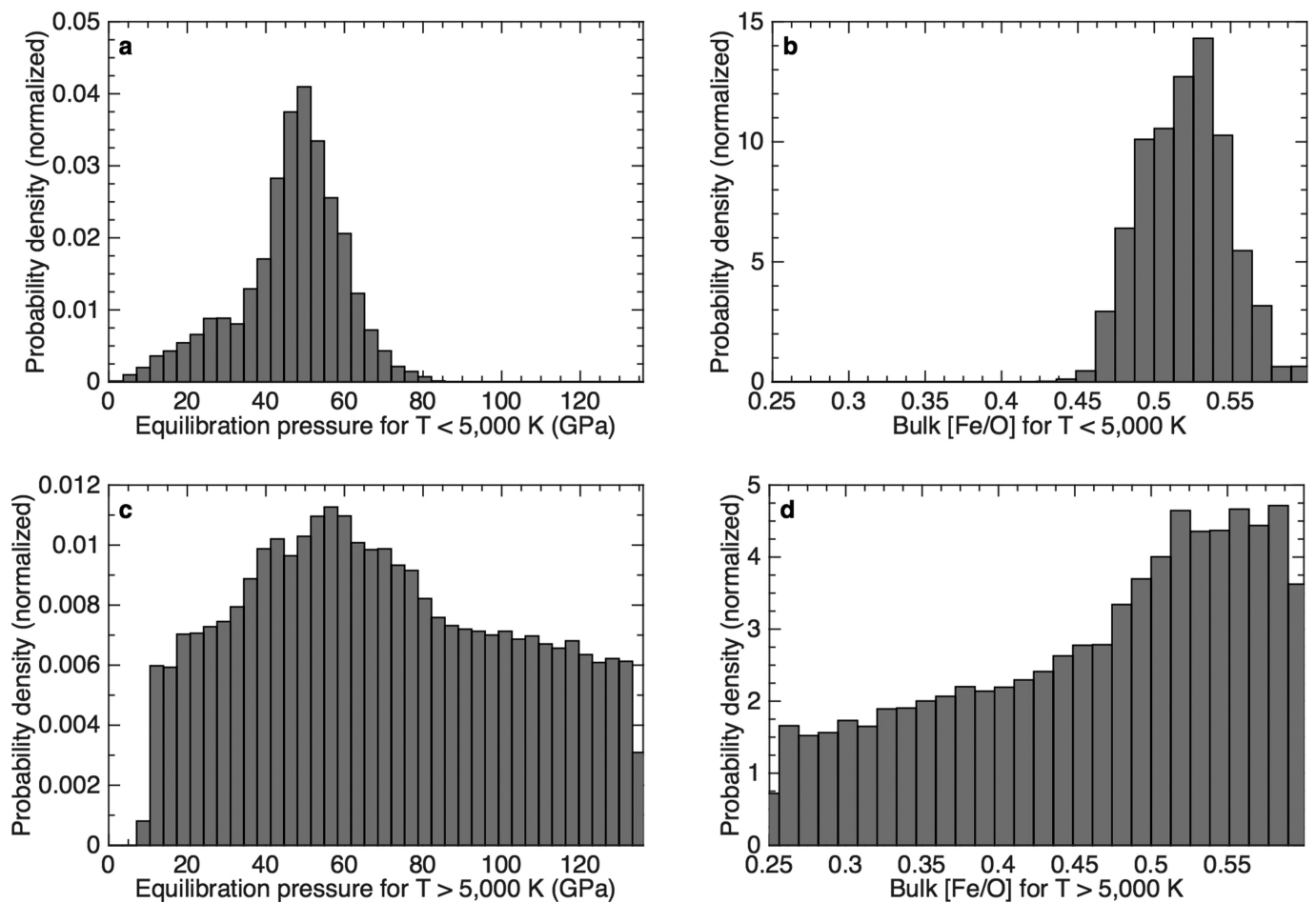
Extended Data Figure 1 | One example of the evolution of the composition of the precipitate. Here the core initially contains 2 wt% Mg, 3 wt% Si and 6 wt% O. Additionally, the constants a_O , b_O and c_O are each

reduced by 0.25σ from their estimated mean values. The actual mineralogy of the precipitate (for example, the amount and composition of perovskite) is not modelled in detail.



Extended Data Figure 2 | Additional results from models of Earth's core-mantle differentiation. Normalized distributions of chi-squared values, $p(\chi^2)$, for both models of core formation (**a**), along with posterior

probability densities for the coefficients a_i , b_i , c_i , and e_j^i for various elements i and j used in the two-stage model to calculate partitioning behaviour (**b-e**) and elemental abundances in bulk Earth (**f**).



Extended Data Figure 3 | Posterior probability densities for parameters in the two-stage model of Earth's core-mantle differentiation.

Extended Data Table 1 | Parameters with uncertainties used in models of Earth's differentiation

<i>i</i>	Bulk Earth	σ	Primitive mantle	σ	Units	<i>n</i>	<i>a_i</i>	σ	<i>b_i</i>	σ	<i>c_i</i>	σ
Fe	31.9	1.3	6.3	0.063	wt%							
Ca	1.71	0.10	2.61	0.21	wt%							
Al	1.59	0.10	2.38	0.19	wt%							
Mg	15.4	0.62	22.17	0.22	wt%	2	0.1	0.1	-10,851	1,000		
Si	16.1	0.48	21.22	0.21	wt%	4	1.3	0.3	-13,500	900		
Ni	1.82	0.13	0.186	0.009	wt%	2	0.46	0.16	2,700	300	-61	6
Co	880	35	102	5.1	ppm	2	0.36	0.15	1,500	300	-33	5
Nb	0.44	0.044	0.595	0.119	ppm	5	2.66	0.11	-14,032		-199	16
Ta	0.025	0.003	0.043	0.002	ppm	5	0.84	0.09	-13,806		-115	13
V	105	6.3	86	4.3	ppm	3	-1.5	0.3	-2,300	700	9	9
Cr	4,700	235	2,520	252	ppm	2	-0.3	0.2	-2,200	600	-5	7

The table shows the oxygen-free composition of bulk Earth²³, elemental abundances in the primitive mantle³², and constants *a_i*, *b_i*, and *c_i* used to model partitioning behaviour^{12,22}. For O, *a* = 0.6 ± 0.4, *b* = -3,800 ± 900, and *c* = 22 ± 14. Likewise, $e_0^V = -0.077 \pm 0.008$, $e_{Si}^V = 0.039 \pm 0.014$, and $e_0^{Cr} = -0.037 \pm 0.007$, which are used to calculate temperature-dependent interaction parameters as described in ref. 22.

Integrative modelling reveals mechanisms linking productivity and plant species richness

James B. Grace¹, T. Michael Anderson², Eric W. Seabloom³, Elizabeth T. Borer³, Peter B. Adler⁴, W. Stanley Harpole^{5,6,7}, Yann Hautier⁸, Helmut Hillebrand⁹, Eric M. Lind³, Meelis Pärtel¹⁰, Jonathan D. Bakker¹¹, Yvonne M. Buckley¹², Michael J. Crawley¹³, Ellen I. Damschen¹⁴, Kendi F. Davies¹⁵, Philip A. Fay¹⁶, Jennifer Firn¹⁷, Daniel S. Gruner¹⁸, Andy Hector¹⁹, Johannes M. H. Knops²⁰, Andrew S. MacDougall²¹, Brett A. Melbourne¹⁵, John W. Morgan²², John L. Orrock¹⁴, Suzanne M. Prober²³ & Melinda D. Smith²⁴

How ecosystem productivity and species richness are interrelated is one of the most debated subjects in the history of ecology¹. Decades of intensive study have yet to discern the actual mechanisms behind observed global patterns^{2,3}. Here, by integrating the predictions from multiple theories into a single model and using data from 1,126 grassland plots spanning five continents, we detect the clear signals of numerous underlying mechanisms linking productivity and richness. We find that an integrative model has substantially higher explanatory power than traditional bivariate analyses. In addition, the specific results unveil several surprising findings that conflict with classical models^{4–7}. These include the isolation of a strong and consistent enhancement of productivity by richness, an effect in striking contrast with superficial data patterns. Also revealed is a consistent importance of competition across the full range of productivity values, in direct conflict with some (but not all) proposed models. The promotion of local richness by macroecological gradients in climatic favourability, generally seen as a competing hypothesis⁸, is also found to be important in our analysis. The results demonstrate that an integrative modelling approach leads to a major advance in our ability to discern the underlying processes operating in ecological systems.

Ecosystem productivity and species diversity are essential to the ability of natural systems to provide goods and services. Yet, for decades there has been debate over their interrelationship. In the 1970s and 1980s, conflicting models predicted that elevated productivity would lead to reductions in species richness^{4–7}. Beginning in the mid-1990s, scientists started to seriously debate another possibility: that richness could promote productivity^{9–12}. While experimental studies generally support a biodiversity enhancement of productivity^{13,14}, the precise strength of the effect in natural systems and the relationship of this process to other factors that can influence productivity remain major questions. Adding to the debate, macroecological theories propose that regional diversity is controlled by gradients in climatic favourability and evolutionary history¹⁵ and that these larger-scale effects are important determinants of smaller-scale diversity patterns⁸.

The search for a canonical bivariate productivity–richness relationship lies at the heart of the debate among ecologists. This pursuit is fuelled, in part, by the history of the discussion, which has focused on bivariate predictions¹⁶. At the same time, it is also seen by some as a means of assessing the overall importance of various mechanisms operating in natural systems. While many different mechanisms have been discussed, the primary competing theories make four main conflicting predictions: (1) richness and productivity should increase together with increasing resources and environmental favourability until limits to coexistence are reached at high productivity and richness declines, producing a humped-shape relationship^{4–7,17}; (2) richness promotes productivity, leading to a positive relationship^{6,9}; (3) richness and productivity increase together because climatic gradients in productivity lead to increased regional species pools, creating a positive relationship but from a separate mechanism⁸; and (4) the richness–productivity relationship will be of inconsistent form because the mechanisms controlling them vary in their scale-dependence and relative importance^{18,19}.

Empirical tests of the generality of hypothesized bivariate productivity–richness patterns have reported a wide variety of results and have produced substantial discussion^{20–23}. Recent global studies^{2,3} have disagreed with regard to whether a coherent pattern exists for natural grasslands. What has been agreed upon, however, is that the low explanatory power coming from conventional analyses suggests the need to pursue an integrative understanding of the causal mechanisms controlling productivity–richness relationships.

One potential explanation for why debate over mechanisms is proving difficult to resolve is because productivity and richness are jointly controlled by a complex network of processes^{1,21,24–28}. Overcoming the challenge of evaluating more complex hypotheses requires both advanced statistical modelling approaches and large-scale systematic data collection efforts. Here, we used structural equation modeling²⁹ to integrate key predictions from competing theories into a multi-process hypothesis for evaluation. We then evaluated the hypothesis using data collected for that purpose by a global consortium, the Nutrient Network (<http://nutnet.org>). The data collected comprise samples

¹US Geological Survey, Wetland and Aquatic Research Center, 700 Cajundome Boulevard, Lafayette, Louisiana 70506, USA. ²Department of Biology, 206 Winston Hall, Wake Forest University, Box 7325 Reynolda Station, Winston-Salem, North Carolina 27109, USA. ³Ecology, Evolution, and Behavior, University of Minnesota, 1987 Upper Buford Circle, St Paul, Minnesota 55108, USA. ⁴Department of Wildland Resources and the Ecology Center, Utah State University, 5230 Old Main, Logan, Utah 84322, USA. ⁵Department of Physiological Diversity, Helmholtz Center for Environmental Research – UFZ, Permoserstrasse 15, 04318 Leipzig, Germany. ⁶German Centre for Integrative Biodiversity Research (iDiv), Deutscher Platz 5e, D-04103 Leipzig, Germany. ⁷Martin Luther University Halle-Wittenberg, Am Kirchtor 1, 06108 Halle (Saale), Germany. ⁸Ecology and Biodiversity Group, Department of Biology, Utrecht University, Padualaan 8, Utrecht 3584 CH, The Netherlands. ⁹Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Schleusenstrasse 1, Wilhelmshaven D-26381, Germany. ¹⁰Institute of Ecology and Earth Sciences, University of Tartu, Lai 40, Tartu 51005, Estonia. ¹¹School of Environmental and Forest Sciences, University of Washington, Box 354115, Seattle, Washington 98195-4115, USA. ¹²School of Natural Sciences, Zoology, Trinity College Dublin, The University of Dublin, Dublin 2, Ireland. ¹³Department of Biological Sciences, Imperial College London, Silwood Park, Ascot, Berkshire SL5 7PY, UK. ¹⁴Department of Zoology, University of Wisconsin, 430 Lincoln Drive, Madison, Wisconsin 53706, USA. ¹⁵Department of Ecology and Evolutionary Biology, UCB 334, University of Colorado, Boulder, Colorado 80309, USA. ¹⁶Grassland Soil and Water Research Laboratory, United States Department of Agriculture Agricultural Research Service, 808 East Blackland Road, Temple, Texas 76702, USA. ¹⁷15 Queensland University of Technology, School of Earth, Environment and Biological Sciences, Brisbane, Queensland 4001, Australia. ¹⁸Department of Entomology, University of Maryland, College Park, 4112 Plant Sciences, College Park, Maryland 20742, USA. ¹⁹Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK. ²⁰School of Biological Sciences, 348 Manter Hall, University of Nebraska, Lincoln, Nebraska 68588, USA. ²¹Department of Integrative Biology, University of Guelph, Guelph, Ontario N1G 2W1, Canada. ²²Department of Ecology, Environment, and Evolution, La Trobe University, Bundoora, Victoria 3083, Australia. ²³CSIRO Land and Water, Private Bag 5, Wembley, Western Australia, 6913, Australia. ²⁴Department of Biology, Colorado State University, 1878 Campus Delivery, Fort Collins, Colorado 80526, USA.

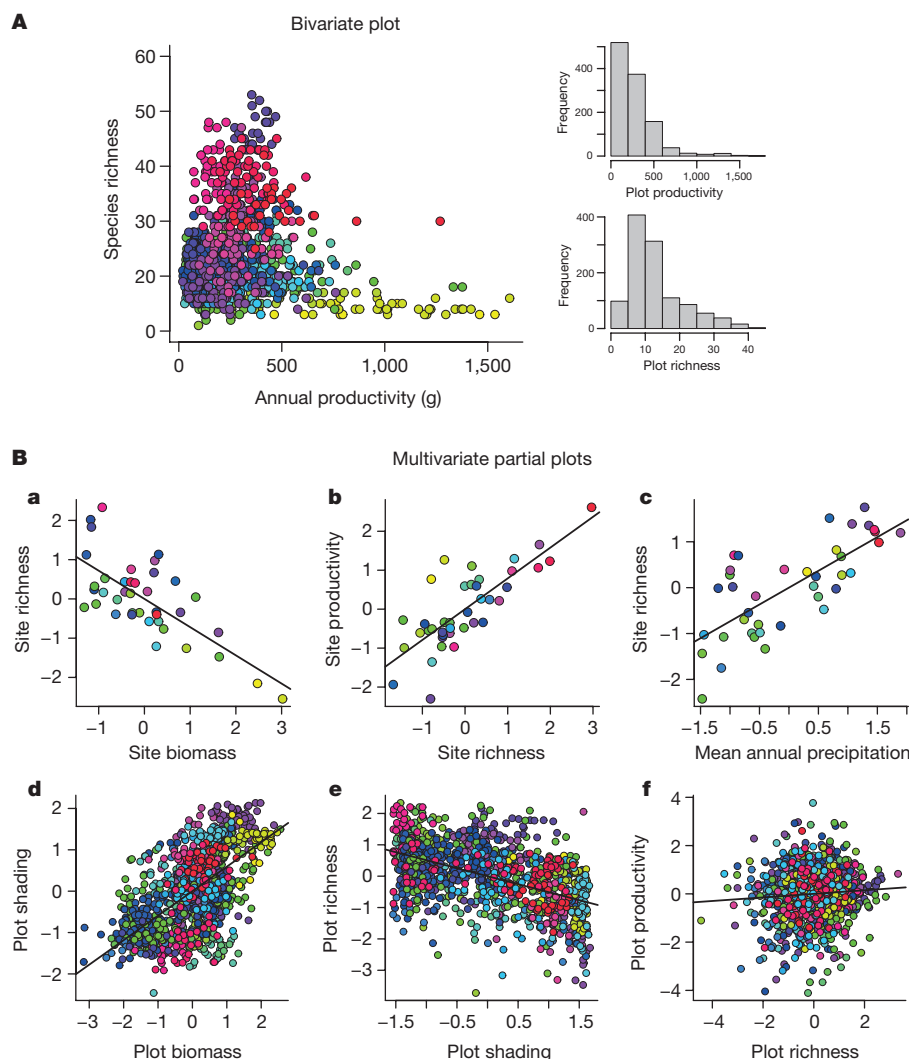


Figure 1 | Comparison between low-dimension (top panel) and high-dimension (bottom panel) examinations of data. **A**, Raw bivariate plot of above-ground productivity and species richness in 1-m² plots ($n = 1,126$). Different sites are represented in the graphs by different colours, assigned by mean site richness from low (yellow) to high (red). **B**, Plots **a–c** visualize site level partial relationships indicated by corresponding letters in Fig. 2 ($n = 39$ sites). Plots **d–f** visualize plot level partial relationships indicated by corresponding letters in Fig. 2 ($n = 1,126$ 1-m² plots). Units are standardized residual deviations from predicted partial scores.

from 1,126 plots collected at 39 grass-dominated sites around the world. Variables measured include plant species richness, productivity (measured as the annual biomass increment), total biomass (the accumulated non-woody biomass, live and dead, including litter), along with many of the drivers hypothesized to be important for regulating their variations. Additional information is provided in the Methods.

To integrate theoretical expectations from competing theories, we mined the productivity–diversity literature to determine the main theoretical constructs discussed and the hypothesized interconnections between constructs (see Methods). We used this information to develop a structural equation meta-model that assimilates the essential theoretical constructs and hypothesized connections into a network of multivariate expectations (Extended Data Fig. 1, Extended Data Table 1 and Supplementary Information). This meta-model, along with the available data, guided our development of a structural equation model for empirical evaluation. We evaluated model–data consistency to determine whether there were missing linkages in the initial model as well as to determine the support for proposed links. We further addressed the question, ‘what dimension of model (that is, number of parameters and linkages) is required to detect the signals in the underlying data?’. For this, we evaluated lower-dimensional versions of the model by removing linkages and re-evaluating against the data. More methodological detail is provided in the Methods and Supplementary Information.

A simple bivariate plot of richness against productivity (Fig. 1A) reveals little about the underlying mechanisms. Previous analyses of such bivariate relations have found it difficult to even detect significant associations^{2,30}. However, our analysis based on an integrative

model reveals strong, clear signals consistent with numerous proposed mechanisms, including several that are not at all suggested from the bivariate data.

First, we found clear evidence that the accumulation of total biomass (hereafter simply ‘biomass’) leads to a negative effect on species richness. At the site level, the partial effect (r_{∂}) of biomass on richness in the model was strong (Figs 1B, a and 2; $r_{\partial} = -0.77$). The reduction of richness was not found to be mediated by our one-time measurement of average shading at the ground surface, which was subsequently dropped from the model. At the plot level, however, we found evidence that biomass increases shading ($r_{\partial} = 0.56$), which in turn, decreases richness (Figs 1B, d, e and 2; $r_{\partial} = -0.34$). The negative effects of biomass on richness appear consistent with long-standing hypotheses that predict a hump-shaped productivity–richness relationship due to competitive dominance at high productivity^{5,17}. However, while those hypotheses assume increasing competitive intensity with increasing productivity, our results reveal a linear effect across the full range of biomass observed in this study (Fig. 1B, a).

Second, we found a positive, linear enhancement of productivity by richness in the model. This effect was among the strongest found at the site-scale (Figs 1B, b and 2; $r_{\partial} = 0.67$), and was detectable, although weak, at the plot-scale (Figs 1B, f and 2; $r_{\partial} = 0.02$). A surprising feature of the site-level result is the apparent absence of a levelling off of the biodiversity enhancement of productivity at higher levels of richness. Such a continuous effect has been theorized for larger-scale studies and contrasts with the asymptotic levelling off usually

Table 1 | Select standardized partial effect sizes (and standard errors) ranked by magnitude and proposed interpretations

Effect	Magnitude	Proposed interpretation
Site-scale		
Soil fertility → productivity	1.104 (0.220)	Productivity variations are most strongly related to spatial variations in site fertility.
Site biomass → richness	−0.771 (0.143)	High biomass sites have depressed species richness, presumably via some form of competitive dominance.
Site richness → productivity	0.671 (0.214)	Increasing richness contributes to higher productivity.
Climate → richness	0.669 (0.113)	Richness increases with increasing mean annual precipitation during warmest quarter.
Heterogeneity → richness	0.627 (0.119)	Species coexistence strongly regulated by within-site spatial heterogeneity in vegetation cover.
Climate → productivity	0.589 (0.195)	Combined effects of macrogradients in temperature and precipitation are moderately important in controlling site-to-site variations in productivity.
Soil suitability → richness	0.439 (0.107)	Richness increases with decreasing sand and silt content of soils.
Disturbance → richness	−0.251 (0.116)	Sites subject to strong anthropogenic alteration (for example, pasturing) notably lower in richness.
Disturbance → biomass	−0.185 (0.084)	Local grazing contributed to some biomass variation.
Correlation between soil fertility and soil suitability	−0.56 (0.109)	Influences of resources and filters on richness are distinct from those that regulate productivity.
Plot-scale		
Biomass → shading	0.559 (0.089)	Shading fairly strongly tied to biomass, although morphology and other features probably also important.
Soil suitability → richness	0.404 (0.074)	Plot-to-plot variations in richness tied to local variations in environmental edaphic filters.
Shading → richness	−0.342 (0.064)	Richness variations within site partially controlled by variations in competition via shading.
Soil suitability → shading	0.249 (0.098)	Variations in community traits driven by soil resources result in differences in amount of shading per gram.
Richness → productivity	0.017 (0.006)	Effects of natural plot-to-plot variations in richness can be discerned as an independent process in this sample.
Soil fertility → productivity	0.0 (NS)	Plot productivity largely determined by site-level productivity, with no significant local control.

Site-scale and plot-scale effects are presented separately.

found in experimental (smaller-scale) studies¹³. Previous attempts to isolate an effect of richness on productivity with observational data using simpler models have failed to do so (see Supplementary Information).

Third, we found strong and independent influences of macroclimate and soils on richness and productivity. The standardized effect sizes provide insights into the relative importance of these processes (Table 1). At the site level, productivity was most strongly related to soil fertility, while richness was most strongly related to climate and soil suitability, with heterogeneity and disturbance also important (Fig. 2). Rather than being made up of similar environmental factors, the soil environmental drivers of richness and productivity were negatively

correlated at the site level (Table 1, $r_{\theta} = -0.56$), supporting previous claims of their semi-independence²¹. Thus, theories that presume a simultaneous increase in productivity and richness with increasing environmental favourability (see Supplementary Information) fail to correspond with the independent responses to environmental drivers observed in natural systems.

Our results show that failure to account for the variation in richness and productivity explained by the environmental drivers would make it difficult to detect the reciprocal influences of productivity and richness on each other. In fact, our capacity to isolate underlying processes was highly sensitive to model dimensionality, where dimensionality refers to the number of measured determinants of productivity and

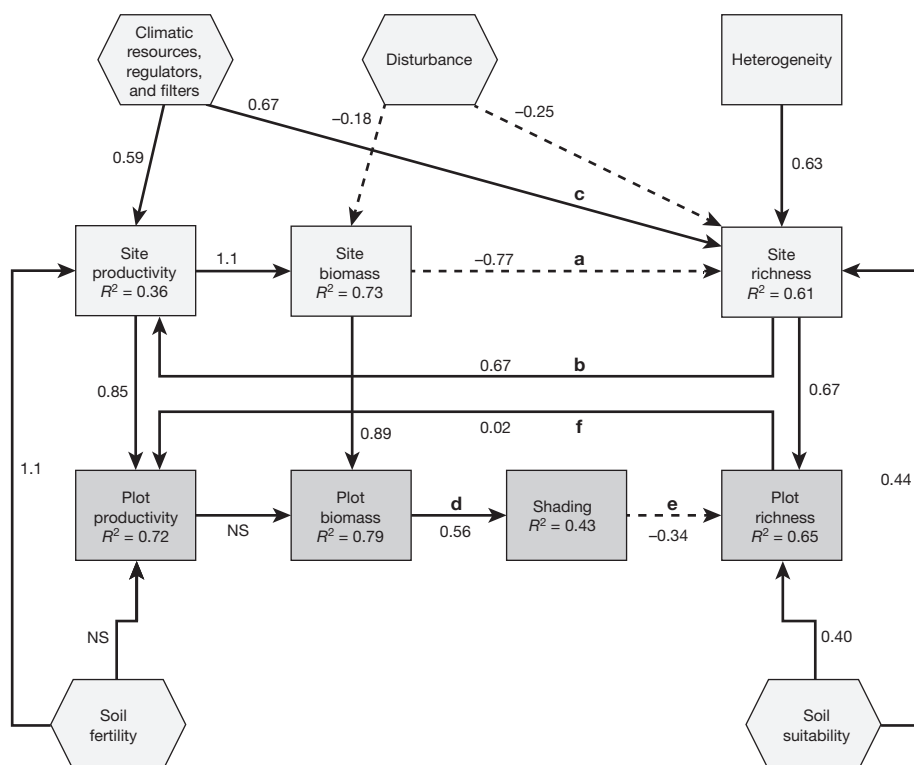


Figure 2 | Structural equation model representing connections between productivity and richness supported by the data. ‘Biomass’ refers to total above-ground accumulated biomass. Letters correspond to partial plots shown in Fig. 1B. Solid arrows represent positive effects, dashed arrows represent negative effects. For the site-level submodel, test statistic = 13.518, with 13 model degrees of freedom and $P = 0.409$ (indicating close model-data fit). For the plot-level submodel, robust test statistic = 21.907, with 16 model degrees of freedom and $P = 0.146$ (again indicating close model-data fit). Relative effect sizes presented in Table 1.

richness included in the model (Extended Data Table 2). At both site and plot levels, models omitting either productivity or biomass (but not both) still permitted us to detect the feedback from richness to biomass production. Any other simplifications at the site-level, however, resulted in a failure to detect previously detected pathways and resulted in a dramatic loss of signal (as indicated by reduced values of R^2 in the model).

Regarding scale dependence, plot-level values of productivity, biomass, and richness were strongly related to site-level estimates (Fig. 2), as is common with hierarchical data. This should be interpreted as meaning much of the overall plot-to-plot variation in productivity, biomass, and richness can be ascribed to site-to-site variations in those properties. In this case, within-site variations in productivity were explained solely by site-level productivity, as there were no predictors for remaining among-plot variations. Within-site variations in richness, however, were additionally explained by within-site variations in soil suitability and shading. Also sensitive to scale was the strength of the feedback from richness to productivity, which was much stronger at the site scale. While multiple factors probably play a role in this scale-dependence, the simplest explanation here may be the smaller span of conditions sampled within sites compared with across sites.

Finally, in contrast to a bivariate model, which our analyses suggest can explain no more than 10% of the observed variation in richness, our structural equation model explains 61% of the variation in richness among sites, and 65% of the variation in richness among plots. An ability to explain a substantial portion of the variation in richness is tremendously important for potential conservation applications. Model complexity is also important because of its more detailed mapping onto nature, as our model can make statements about how both specific management actions (such as reduction of biomass through mowing or increase in soil fertility through fertilization), as well as shifts in climate conditions, may alter both productivity and species richness.

Our findings give reason for optimism about the future of ecology as a more precise and less ambiguous science. We show that many of the proposed processes connecting productivity and richness offered during previous decades operate simultaneously as parts of a whole system of effects. Details of the findings, however, refine many of our assumptions about how those processes operate. Our field's previous failure to resolve debate about productivity–richness relationships stems from a lack of integration of ideas and absence of simultaneous tests of their combined implications. By integrating and testing those ideas, our approach provides a systems-level understanding and improves our chances to foresee the possible consequences of human alteration of environmental factors, productivity, and richness now occurring worldwide.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 September; accepted 8 December 2015.

Published online 13 January 2016.

- Willig, M. R. Biodiversity and productivity. *Science* **333**, 1709–1710 (2011).
- Adler, P. B. *et al.* Productivity is a poor predictor of plant species richness. *Science* **333**, 1750–1753 (2011).
- Fraser, L. H. *et al.* Plant ecology. Worldwide evidence of a unimodal relationship between productivity and plant species richness. *Science* **349**, 302–305 (2015).
- Grime, J. P. Competitive exclusion in herbaceous vegetation. *Nature* **242**, 344–347 (1973).
- Huston, M. A. A general hypothesis of species diversity. *Am. Nat.* **113**, 81–101 (1979).
- Tilman, D. *Resource Competition and Community Structure* (Princeton Univ. Press, 1982).
- Taylor, D. R., Aarssen, L. W. & Loehle, C. On the relationship between r/K selection and environmental carrying capacity: a new habitat template for plant life history strategies. *Oikos* **58**, 239–250 (1990).

- Gillman, L. N. & Wright, S. D. The influence of productivity on the species richness of plants: a critical assessment. *Ecology* **87**, 1234–1243 (2006).
- Naeem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H. & Woodfin, R. M. Declining biodiversity can alter the performance of ecosystems. *Nature* **368**, 734–737 (1994).
- Tilman, D., Wedin, D. & Knops, J. Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature* **379**, 718–720 (1996).
- Huston, M. A. Hidden treatments in ecological experiments: re-evaluating the ecosystem function of biodiversity. *Oecologia* **110**, 449–460 (1997).
- Grime, J. P. Biodiversity and ecosystem function: the debate deepens. *Science* **277**, 1260–1261 (1997).
- Cardinale, B. J. *et al.* Biodiversity loss and its impact on humanity. *Nature* **486**, 59–67 (2012).
- Tilman, D., Reich, P. B. & Isbell, F. Biodiversity impacts ecosystem productivity as much as resources, disturbance, or herbivory. *Proc. Natl Acad. Sci. USA* **109**, 10394–10397 (2012).
- Hawkins, B. A. *et al.* Energy, water, and broad-scale geographic patterns of species richness. *Ecology* **84**, 3105–3117 (2003).
- Pärtel, M., Zobel, K., Laanisto, L., Szava-Kovats, R. & Zobel, M. The productivity–diversity relationship: varying aims and approaches. *Ecology* **91**, 2565–2567 (2010).
- Grime, J. P. *Plant Strategies and Vegetation Processes* (John Wiley, 1979).
- Chase, J. M. & Leibold, M. A. Spatial scale dictates the productivity–biodiversity relationship. *Nature* **416**, 427–430 (2002).
- Scheiner, S. M. & Jones, S. Diversity, productivity and scale in Wisconsin vegetation. *Evol. Ecol. Res.* **4**, 1097–1117 (2002).
- Waide, R. *et al.* The relationship between productivity and species richness. *Annu. Rev. Ecol. Syst.* **30**, 257–300 (1999).
- Grace, J. B. The factors controlling species density in herbaceous plant communities: an assessment. *Perspect. Plant Ecol. Evol. Syst.* **2**, 1–28 (1999).
- Mittelbach, G. G. *et al.* What is the observed relationship between species richness and productivity? *Ecology* **82**, 2381–2396 (2001).
- Whittaker, R. J. Meta-analyses and mega-mistakes: calling time on meta-analysis of the species richness–productivity relationship. *Ecology* **91**, 2522–2533 (2010).
- Borer, E. T. *et al.* Herbivores and nutrients control grassland plant diversity via light limitation. *Nature* **508**, 517–520 (2014).
- Schmid, B. The species richness–productivity controversy. *Trends Ecol. Evol.* **17**, 113–114 (2002).
- Loreau, M., Naeem, S. & Inchausti, P. *Biodiversity and Ecosystem Functioning: Synthesis and Perspectives* (Oxford Univ. Press, 2002).
- Cardinale, B. J., Bennett, D. M., Nelson, C. E. & Gross, K. Does productivity drive diversity or vice versa? A test of the multivariate productivity–diversity hypothesis in streams. *Ecology* **90**, 1227–1241 (2009).
- Scherber, C. *et al.* Bottom-up effects of plant diversity on multitrophic interactions in a biodiversity experiment. *Nature* **468**, 553–556 (2010).
- Grace, J. B. *et al.* Guidelines for a graph-theoretic implementation of structural equation modeling. *Ecosphere* **3**, art73 (2012).
- Grace, J. B., Adler, P. B., Stanley Harpole, W., Borer, E. T. & Seabloom, E. W. Causal networks clarify productivity–richness interrelations, bivariate plots do not. *Funct. Ecol.* **28**, 787–798 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements J.B.G. was supported by the US Geological Survey Ecosystems and Climate and Land use Change Programs. This work uses data from the Nutrient Network (<http://nutnet.org>) experiment, funded at the site scale by individual researchers. Coordination and data management were supported by funding to E.T.B. and E.W.S. from the National Science Foundation (NSF) Research Coordination Network (NSF-DEB-1042132) and Long Term Ecological Research (NSF-DEB-1234162 to Cedar Creek LTER) programs and the UMN Institute on the Environment (DG-0001-13). The Minnesota Supercomputer Institute hosts project data. The use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US Government. Support for site-level activities is acknowledged in the Supplementary Information. We thank D. Laughlin for comments on the manuscript.

Author Contributions E.W.S., E.T.B., W.S.H. and E.M.L. are Nutrient Network coordinators. J.B.G. and T.M.A. developed and framed the research questions. T.M.A., E.W.S., E.T.B., P.B.A., W.S.H., Y.H., H.H., J.D.B., Y.M.B., M.J.C., E.I.D., K.F.D., P.A.F., J.F., D.S.G., A.H., J.M.H.K., A.S.M., B.A.M., J.W.M., J.L.O., S.M.P. and M.D.S. collected data used in this analysis. T.M.A. assembled the data and performed initial analyses. J.B.G. analysed the data and wrote the paper with contributions and input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.B.G. (gracej@usgs.gov).

METHODS

Development of meta-model hypothesis. A review and accounting of the history of claims and disputed points in the published literature was developed before construction of the meta-model that guided this analysis (Extended Data Fig. 1 and Supplementary Information). During this review, attention was paid to the theoretical constructs invoked by various authors, since our goal was to provide a framework that had the potential to clarify and resolve disputed points. Attention was also paid to types of variable measured by different authors, as the relationship between constructs and measurements constitutes one of the several sources of ambiguity and confusion^{31,32}. An in-depth description of the literature synthesized to generate the meta-model is presented in the Supplementary Information.

Data. Data collected by the Nutrient Network Cooperative³³ was used to design and evaluate a structural equation model based on the meta-model presented. The Nutrient Network is a distributed, coordinated research cooperative. Sites in the Network are dominated primarily by herbaceous vegetation and intended to represent natural/semi-natural grasslands and related ecosystems worldwide. Individual sites were selected to accommodate at least a 1,000 m² study design footprint. Most sites sampled vegetation in 2007, although 12 sites sampled in 2008 or 2009. No statistical methods were used to predetermine sample size. Samples were collected using a completely randomized block design. The standard design has three blocks and ten plots per block at each site, although some sites deviate slightly from this design. A few sites are grazed or burned before sampling, consistent with their traditional management. Further details on site selection and design can be found at http://www.nutnet.org/exp_protocol.

In this study, we analysed data from 39 of the 45 sites considered in ref. 2 possessing a complete set of covariates (Extended Data Table 3). While ref. 2 only examined bivariate relations between productivity and richness, our analyses brought in many additional variables (Extended Data Table 1) so that we could address the many hypotheses embodied in the meta-model. Individual plots with greater than 10% woody plant cover were omitted from consideration to maintain comparability in total biomass across plots. This step resulted in the removal of 73 plots, leaving 1,126 plots in the data set analysed. Four plots were omitted owing to incomplete plant data and one for incomplete light data. For two of the sites, live mass was estimated from total mass using available information on the proportion of live to total. One apparent measurement error was detected for light data and the associated plot removed from the analysed sample. Random imputation methods³⁴ were used for cases where there were missing soil measurements at a site. The decision to use this approach was based on weighing the demerits of deleting nearly complete multivariate data records versus introducing a modest amount of random error through the imputation process.

Study plots in this investigation had a perimeter of 5 m × 5 m and were separated by 1 m walkways. A single 1 m × 1 m subplot within each plot was permanently marked and sampled for species richness during the season of peak biomass. Sites with strong seasonal variation in composition were sampled twice during the season to assemble a complete list of species. To obtain an estimate of site-level richness, we used a jack-knife procedure³⁵. (Because there have been some recent advances in the reduction of certain sources of bias in richness estimation³⁶, we checked our original results by computing site-level richness using the new iNEXT R package. The correlation between the two estimates of richness was found to be 0.972.)

Productivity and total above-ground biomass were sampled immediately adjacent to the permanent vegetation subplot. Vegetation was sampled destructively by clipping at ground level all above-ground biomass of individual plants rooted within two 0.1-m² (10 cm × 100 cm) strips. Harvested plant material was sorted into the current year's live and recently senescent material, and into previous year's growth (including litter). For shrubs and sub-shrubs, the current year's leaves and stems were collected. Plant material was dried at 60 °C to a constant mass and weighed to the nearest 0.01 g. We used the current year's biomass increment as our estimate of annual above-ground productivity, which commonly serves as a measurable surrogate for total productivity^{37,38}. All sites used this protocol to estimate productivity (except for the Sevilleta, New Mexico, site which relied on species-specific allometric relationships³⁹). Total above-ground biomass was computed as the sum of the current year's biomass and that from previous years and included remaining dead material (litter). Photosynthetically active radiation was measured at the time of peak biomass, both above the vegetation and at the ground surface, the ratio representing the proportion of available light reaching the ground. Degree of shading was computed as 1.0 minus the proportion of light reaching the ground.

Within each plot, 250 g of soil were collected and air dried for processing and soil archiving. Total soil %C and %N were measured using dry combustion gas chromatography analysis (COSTECH ESC 4010 Element Analyzer) at the University of Nebraska. All other soil analyses were performed at A&L Analytical

Laboratory, Memphis, Tennessee, USA; these included the following: extractable soil phosphorus and potassium were quantified using the Mehlich-3 extraction method, and parts per million concentration estimated using inductively coupled plasma-emission spectrometry. Soil pH was quantified with a pH probe (Fisher Scientific) in a slurry made from 10 g dry soil and 25 ml of deionized water. Soil texture, expressed as the percentage sand, percentage silt, and percentage clay, was measured on 100 g dry soil using the Buoycous method. Further details on sampling methodology are at http://www.nutnet.org/exp_protocol.

Climatic characteristics were obtained for each site from version 1.4 of BioClim, which is part of the WorldClim⁴⁰ set of global climate layers at 1 km² spatial resolution. To represent measures of temperature and precipitation with meaningful relationships to plant growth in global grasslands, we selected mean temperature of the wettest quarter of the year (BIO8) and total precipitation of the warmest quarter of the year (BIO18). Climate values were extracted using universal transverse Mercator (UTM) coordinates collected near the centre of each site.

Several derived variables were developed to include in the modelling effort. To represent within-site heterogeneity, coefficients of variation were computed for the site-level model based on plot-to-plot variation in plot-level measures. This allowed us to examine the explanatory value of heterogeneity in soil nitrogen, phosphorus, potassium, and pH, as well as heterogeneity in biomass and light interception. Indices of total resource supply and resource imbalance were also calculated using the method of ref. 27 and evaluated for inclusion in our models.

Disturbance history information for the sites was converted into four binary (0,1) variables for analyses; information available included pretreatment history of (1) substantial anthropogenic alteration (for example, conversion to pasture), (2) grazing history, by wild or domestic animals, (3) active management (typically haying or mowing), and (4) fire. Current levels of herbivory were estimated by comparing biomass inside and outside enclosure plots located at each site.

Certain variables were constructed within the structural equation modeling process using the composite index development methods of ref. 41. Consideration of the ideas conveyed by the meta-model (Extended Data Fig. 1) and the specific situation being modelled suggested the need to develop index variables for soil fertility and soil suitability. Soil fertility indices were developed using all measured soil properties and were operationally defined as the drivers of productivity, controlling for all other effects on productivity in the model. Two indices were developed, one for site-to-site variations and another for plot-to-plot variations. Similarly, soil suitability indices were developed for the site- and plot-level data using all measured soil properties as potential contributors and operationally defined as the drivers of richness, controlling for all other effects on richness in the model.

Modelling with composites in structural equation models involved a two-step process. First, we constructed a fully specified structural equation model (as represented in Fig. 2), but providing a specific set of soil properties to serve as formative indicators for soil fertility and soil suitability. Variables that did not contribute to the total model (on the basis of model fit indices) were eliminated individually for the two composites being formed. The resulting prediction equations were used to compute index scores. Then, the model was reconstructed, substituting the indices in place of the collection of individual soil properties. Documentation of this process is provided in the Supplementary Information computer code (R script).

Analyses. A structural equation model was developed based on the ideas embodied in the meta-model, available data, and the principles and procedures laid out in ref. 42. Indicators for constructs were chosen from the set of variables available and quantities that could be computed from them (Extended Data Table 1). The modelling approach used was semi-exploratory in that while we worked to address the general hypothesis embodied in the meta-model, the precise variables (for example, mean annual precipitation versus mean annual precipitation in the warmest quarter of the year) to use for certain constructs (specifically, resource supplies and regulators) were determined empirically. Compositing techniques were used to estimate construct-level effects⁴¹. For comparative purposes, we analysed the bivariate pattern in Fig. 1A using a variety of regression models, including Ricker-type nonlinear models as well as second- and third-order polynomials. A three-parameter Ricker-type model provided the best fit for the data.

Data were screened for distributional properties and nonlinear relations. Several variables were log-transformed as a result of evaluations (Extended Data Table 1). We used the R software platform⁴³ and the lavaan package⁴⁴ along with the lavaan.survey⁴⁵ package for our structural equation model analyses. For the plot-scale model, robust χ^2 tests, as implemented in the lavaan.survey package, were used to judge variable inclusion and model adequacy because of the nested nature of the plot-level data. Each link in the final model was evaluated for significant contribution to the model. Final model fit to data was very good for both submodels. Model fit indices were supplemented by using additional diagnostic evaluations that involve visualizing residual relationships

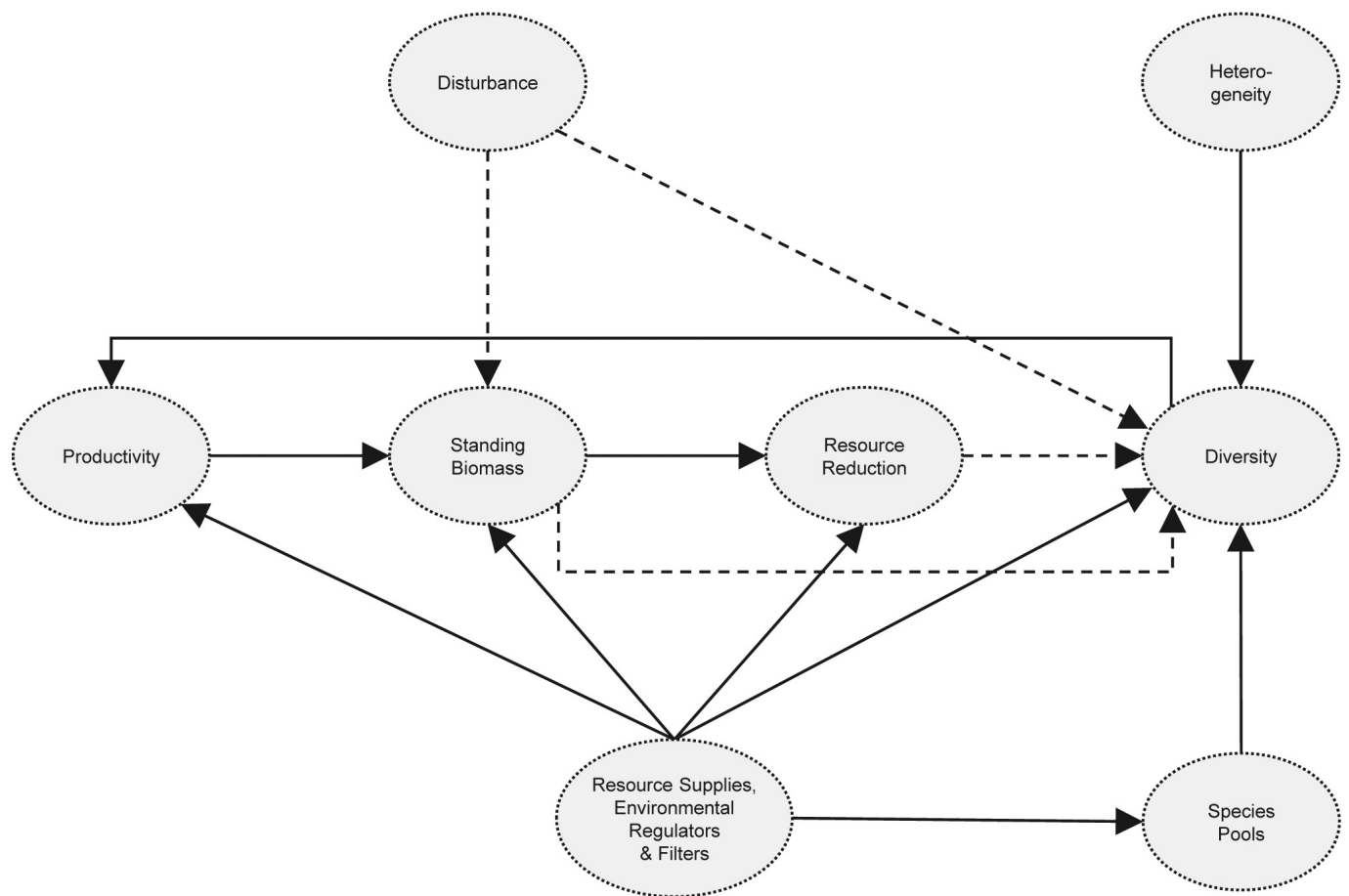
to evaluate conditional independence²⁹. These residual visualizations allowed, among other things, an ability to evaluate linearity assumptions and implement curve-fitting procedures if needed (which was only the case for the composite relationships in this case). Our structural equation model in this case is non-recursive and includes a causal loop. Models of this form are commonplace in structural equation model applications, although they come with some additional assumptions and requirements. Specifically, there is a requirement for unique predictors for the elements involved in loops, a requirement that was met in this case. Additional analysis details are documented in the R script used for the analysis (Supplementary Information).

Multi-level relations were incorporated into the architecture of our model. Several ways to incorporate both site- and plot-level variations in the model were considered and multiple approaches evaluated to ensure results are general. In the model form presented, we chose to follow modern hierarchical modelling principles and allow plot-level observations to depend on site-level parameters, since plots were nested within sites. The result of choosing this approach means site-level explanatory effects can filter down to the plot level while plot-level explanatory variables (for example, pathways from edaphic conditions to plot richness) explain additional plot-to-plot variations in responses that are not predicted from site-level (mean) conditions. Consistent with the capabilities of the structural equation model software used in our analyses (described below), we estimated site- and plot-level submodels using a two-stage approach, first estimating parameters for the site-level component and then using site productivity, biomass, and richness as exogenous predictors in the plot-level component. Comparisons with results from separate site- and plot-level models led to very similar conclusions, although the hierarchical approach used allowed a better integration of processes and greater variance explanation.

One of our objectives in this study was to assess the model dimensionality needed to detect the hypothesized signals in the data. To do this, we started with the most complete model (Fig. 2) and eliminated variables from the model (always retaining richness and some measure of biomass production, either productivity or total biomass). We then made any modifications needed to ensure adequate model-data fit for these reduced-form models. The consequences of model simplification was judged on the basis of signal retention, in particular a loss of capacity to detect signals associated with the remaining parts of the model.

Code availability. The computer script associated with the analyses in this paper is available as part of the Supplementary Information.

31. Grace, J. B., Anderson, T. M., Olff, H. & Scheiner, S. M. On the specification of structural equation models for ecological systems. *Ecol. Monogr.* **80**, 67–87 (2010).
32. Regan, H. M., Colyvan, M. & Burgman, M. A. A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecol. Appl.* **12**, 618–628 (2002).
33. Borer, E. T. *et al.* Finding generality in ecology: a model for globally distributed experiments. *Methods Ecol. Evol.* **5**, 65–73 (2014).
34. Gelman, A. & Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge Univ. Press, 2007).
35. Chao, A. Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**, 783–791 (1987).
36. Chao, A. & Jost, L. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* **93**, 2533–2547 (2012).
37. Lauenroth, W., Hunt, H., Swift, D. & Singh, J. Estimating aboveground net primary production in grasslands: a simulation approach. *Ecol. Modell.* **33**, 297–314 (1986).
38. Oosterheld, M. & McNaughton, S. J. in *Methods in Ecosystem Science* (eds Sala, O. E., Jackson, R. B., Mooney, H. A. & Howarth, R. W.) Ch. 10, 151–157 (Springer, 2000).
39. Muldavin, E. H., Moore, D. I., Collins, S. L., Wetherill, K. R. & Lightfoot, D. C. Aboveground net primary production dynamics in a northern Chihuahuan Desert ecosystem. *Oecologia* **155**, 123–132 (2008).
40. Hijmans, R., Cameron, S., Parra, J., Jones, P. & Jarvis, A. WorldClim, version 1.3 (Univ. California, Berkeley, 2005).
41. Grace, J. B. & Bollen, K. A. Representing general theoretical concepts in structural equation models: the role of composite variables. *Environ. Ecol. Stat.* **15**, 191–213 (2008).
42. Grace, J. B., Scheiner, S. M. & Schoolmaster, D. R. Jr in *Ecological Statistics* (eds Fox, G. A., Negrete-Yankelevich, S. & Sosa, V. J.) Ch. 8, 168–199 (Oxford Univ. Press, 2015).
43. R Development Core Team. R: a language and environment for statistical computing (R Foundation for Statistical Computing, 2012).
44. Rosseel, Y., Oberski, D., Byrnes, J., Vanbrabant, L. & Savalei, V. *lavaan: latent variable analysis (software)* (R Foundation for Statistical Computing, 2013).
45. Oberski, D. *lavaan.survey: an R package for complex survey analysis of structural equation models*. *J. Stat. Softw.* **57**, 1–27 (2014).



Extended Data Figure 1 | Structural equation meta-model showing hypothesized probabilistic expectations based on literature related to the productivity-diversity debate. Solid lines represent expected positive effects, dashed lines represent expected negative effects.

Literature and meta-model development are discussed in the Supplementary Information. Specific implementations of this generalized model for particular cases will probably differ in detail as appropriate for the situation and available data.

Extended Data Table 1 | Model variables and their indicators*

Model Variables	Indicator Variables	Units**
Site Richness (log)	estimated site richness	number/site
Site Biomass (log)	peak season total above-ground biomass, including litter	mean g/m ²
Site Productivity (log)	peak season live above-ground biomass increment	mean g/m ²
Plot Richness (log)	species in a plot	number/plot
Shading (log)	proportional reduction in light at ground surface	0.0 – 1.0
Plot Biomass (log)	peak season total above-ground biomass, including litter	g/m ² per plot
Plot Productivity (log)	peak season live above-ground biomass increment	g/m ² per plot
Climate Effect on Site Richness	mean precipitation in warmest quarter	mm
Climate Effect on Site Productivity	mean precipitation in warmest quarter temperature in wettest quarter	mm, °C
Disturbance Effect on Site Richness	history of major anthropogenic influences	0 or 1
Disturbance Effect on Site Biomass	herbivory (based on exclosure studies)	proportion
Heterogeneity	CV for variations in vegetation density among plots at a site (expressed in terms of canopy light interception variations)	unitless
Soil Suitability (site and plot-level)	<i>function of</i> (soil N, P, C, texture, pH) that maximizes richness, controlling for other conditions.	unitless
Soil Fertility (site and plot-level)	<i>function of</i> (soil N, P, C, texture, pH) that maximizes productivity, controlling for other conditions.	unitless

*The data are provided along with the Supplementary Information.

**Units given are for the raw (untransformed) variables.

Extended Data Table 2 | Results of model dimensionality evaluations

Model Level	Model	Signals Lost	R^2 for Richness
Site-level	full model		0.62
	heterogeneity node eliminated	signal from disturbance to richness lost	0.33
	disturbance node eliminated	all paths retained	0.49
	climate node eliminated	signal from disturbance to richness lost; signal from richness to productivity lost	0.37
	soil fertility and suitability eliminated	signal from disturbance to richness lost; signal from richness to productivity lost; signal from climate to productivity lost	0.16
	biomass node eliminated	all signals were retained	0.50
	eliminate path from richness to productivity	significant misspecification, model fails to fit data unless path reinstated	
Plot-level	full model		0.67
	soil fertility and suitability eliminated	all signals were retained	0.51
	shading node eliminated	all signals were retained	0.56
	biomass node eliminated	no substantive signals lost	0.65
	productivity node eliminated	all signals were retained	0.67
	eliminate path from richness to productivity	significant misspecification, model fails to fit data unless path reinstated	

Models of different complexity were evaluated to determine the potential for model simplification. The bases for comparison were 'full models' for each site, as shown in Fig. 2. The consequences of removing various components of the models are summarized under the columns 'Signals lost' and ' R^2 for richness'.

Extended Data Table 3 | Basic information on the study sites included in the final analyses

Site Code	Site Name	Continent	Region	Country	State	Site Lead PIs
amcamp.us	American Camp	North America	Pacific Coast (North America)	US	Washington	Jonathan D. Bakker; Janneke Hille Ris Lambers
azi.cn	Azi	Asia	Asia	CN	Gansu	Chengjin Chu; Qi Li; Wei Li; Gang Wen; Guozhen Du
barta.us	Barta Brothers	North America	Central plains (North America)	US	Nebraska	David Wedin
bldr.us	Boulder South Campus	North America	Central plains (North America)	US	Colorado	Kendi Davies; Brett Melbourne
bnch.us	Bunchgrass (Andrews LTER)	North America	Montane West (North America)	US	Oregon	Eric Seabloom; Elizabeth Borer
bogong.au	Bogong	Australia	Australia	AU	Victoria	John Morgan; Joslin L. Moore
bttr.us	Buttercup (Andrews LTER)	North America	Montane West (North America)	US	Oregon	Eric Seabloom; Elizabeth Borer
burrawan.au	Burrawan	Australia	Australia	AU	Queensland	Jennifer Firm; Yvonne Buckley
cbgb.us	Chichaqua Bottoms	North America	Central plains (North America)	US	Iowa	W. Stanley Harpole; Lori A. Biederman; Kirsten S. Hofmocker; Lauren Sullivan
cdpt.us	Cedar Point Biological Station	North America	Central plains (North America)	US	Nebraska	Johannes M. H. Knops
derr.au	Derrimut	Australia	Australia	AU	Victoria	John Morgan
elliott.us	Elliott Chaparral	North America	Pacific Coast (North America)	US	California	Elsa Cleland
fnly.us	Finley NWR	North America	Pacific Coast (North America)	US	Oregon	Eric Seabloom; Elizabeth Borer
gilb.za	Mt Gilboa	Africa	Sub-saharan Africa	ZA	KwaZulu-Natal	Peter D. Wragg
glac.us	Glacial Heritage	North America	Pacific Coast (North America)	US	Washington	Jonathan D. Bakker; Janneke Hille Ris Lambers
hall.us	Hall's Prairie	North America	Central plains (North America)	US	Kentucky	Rebecca L. McCulley; Jim Nelson
hart.us	Hart Mountain	North America	Montane West (North America)	US	Oregon	David Pyke; Nicole M. DeCrappeo
hnvr.us	Hanover	North America	Atlantic Coast (North America)	US	New Hampshire	Elizabeth M. Wolkovich; Kathryn L. Cottingham
kiny.au	Kinypanial	Australia	Australia	AU	Victoria	John Morgan
look.us	Lookout (Andrews LTER)	North America	Montane West (North America)	US	Oregon	Eric Seabloom; Elizabeth Borer
mtca.au	Mt. Caroline	Australia	Australia	AU	Western Australia	Suzanne M Prober
pape.de	Papenburg	Europe	Europe	DE	Lower Saxony	Helmut Hillebrand
sage.us	Sagehen Creek UCNRS	North America	Montane West (North America)	US	California	Daniel S. Gruner; Louie Yang
sava.us	Savannah River	North America	Atlantic Coast (North America)	US	South Carolina	John L. Orrock; Ellen I. Damschen; Lars Brudvig
sedg.us	Sedgwick Reserve UCNRS	North America	Pacific Coast (North America)	US	California	Carla M D'Antonio; W. Stanley Harpole; Elizabeth Borer; Eric Seabloom
sereng.tz	Serengeti	Africa	Sub-saharan Africa	TZ		T. Michael Anderson
sevi.us	Seville LTER	North America	Central plains (North America)	US	New Mexico	Scott Collins; Laura Ladwig
sgs.us	Shortgrass Steppe LTER	North America	Central plains (North America)	US	Colorado	Cynthia S. Brown; Julia A. Klein; Dana M. Blumenthal; Alan Knapp
shps.us	Sheep Experimental Station	North America	Montane West (North America)	US	Idaho	Peter Adler
sier.us	Sierra Foothills REC	North America	Pacific Coast (North America)	US	California	Eric Seabloom; Elizabeth Borer; W. Stanley Harpole
smith.us	Smith Prairie	North America	Pacific Coast (North America)	US	Washington	Jonathan D. Bakker; Janneke Hille Ris Lambers
spin.us	Spindletop	North America	Central plains (North America)	US	Kentucky	Rebecca L. McCulley; Jim Nelson
summ.za	Summerveld	Africa	Sub-saharan Africa	ZA	KwaZulu-Natal	Peter D. Wragg
temple.us	Temple	North America	Central plains (North America)	US	Texas	Philip A Fay
trel.us	Trelease	North America	Central plains (North America)	US	Illinois	Andrew Leakey; Xiaohui Feng
tyso.us	Tyson	North America	Central plains (North America)	US	Missouri	John L. Orrock; Ellen I. Damschen; Tiffany Knight
ukul.za	Ukulinga	Africa	Sub-saharan Africa	ZA	KwaZulu-Natal	Nicole Hagenah; Kevin P Kirkman
unc.us	Duke Forest	North America	Atlantic Coast (North America)	US	North Carolina	Charles Mitchell; Justin Wright
valm.ch	Val Mustair	Europe	Europe	CH	Grisons	Anita C. Risch; Martin Schuetz

A total of 39 sites from the Nutrient Network (<http://nutnet.org>) possessed sufficiently complete multivariate data to be incorporated into this analysis.

Inter-group violence among early Holocene hunter-gatherers of West Turkana, Kenya

M. Mirazón Lahr^{1,2}, F. Rivera^{1*}, R. K. Power^{1*}, A. Mounier¹, B. Copsey¹, F. Crivellaro¹, J. E. Edung³, J. M. Maillo Fernandez⁴, C. Kiarie², J. Lawrence¹, A. Leakey², E. Mbua⁵, H. Miller¹, A. Muigai⁶, D. M. Mukhongo¹, A. Van Baelen¹, R. Wood⁷, J.-L. Schwenninger⁸, R. Grün^{7,9}, H. Achyuthan¹⁰, A. Wilshaw¹ & R. A. Foley^{1,2}

The nature of inter-group relations among prehistoric hunter-gatherers remains disputed, with arguments in favour and against the existence of warfare before the development of sedentary societies^{1,2}. Here we report on a case of inter-group violence towards a group of hunter-gatherers from Nataruk, west of Lake Turkana, which during the late Pleistocene/early Holocene period extended about 30 km beyond its present-day shore³. Ten of the twelve articulated skeletons found at Nataruk show evidence of having died violently at the edge of a lagoon, into which some of the bodies fell. The remains from Nataruk are unique, preserved by the particular conditions of the lagoon with no evidence of deliberate burial. They offer a rare glimpse into the life and death of past foraging people, and evidence that warfare was part of the repertoire of inter-group relations among prehistoric hunter-gatherers.

The origins of war are controversial. Although it is clear that inter-group violence, including intentional lethal attacks on individuals, is part of the behavioural repertoire of chimpanzees^{4,5}, evolutionary explanations for human violence have been disputed^{1,2}. This uncertainty arises because evidence that informs on the nature of relationships among groups in the past is scarce, and most models of prehistoric inter-group relations rely on ethnographic information from small-scale societies. This information is very variable, partly because of differences in the definition of war, partly because the circumstances of modern hunter-gatherers are not analogous to the past, and partly because of the inclusion of data on both intra- and inter-societal warfare^{1,2,6–8}. After numerous analyses of the scarce ethnographic data, researchers remain deeply divided as to whether antagonistic relations formed a significant element of social life in prehistory^{2,6,7,9–18}. Antagonistic relations can be described as acts of aggression towards an individual or a small foraging/scouting party, or as warfare. Prehistoric cases of the former are difficult to differentiate from inter-personal violence⁶, the most frequent form of lethal aggression among recent foragers²; cases of large-scale violent encounters of two groups are relatively common among settled societies^{6,19}. It is evidence for inter-group violence among prehistoric hunter-gatherers, however, that is extremely rare. The Qadan graveyard at Jebel Sahaba, Sudan, where 23 of 58 bodies show evidence of violence²⁰, stands as the best example. Although undated, the Jebel Sahaba remains were estimated to have a late Pleistocene age (often quoted as 14,000–12,000 years) on the basis of the character of the lithic industry²⁰. The Jebel Sahaba individuals were buried, individually or in small groups, presumably by their own community, after the raids or feuds during which they died. The existence of such cemetery space suggests a level of sedentism that would align the Jebel Sahaba violent deaths to later examples. In contrast, the

human remains from the site of Nataruk record the intentional killing of a small band of foragers, and thus unique evidence of a warfare event among hunter-gatherers in prehistory.

Nataruk is located near the reconstructed margin of the late Pleistocene/early Holocene lake Palaeo-Turkana, at the eastern edge of a small depression that would have formed a lagoon during periods of high precipitation (Extended Data Fig. 1a). In 2012, the remains of a minimum of 27 individuals were discovered partly or completely exposed on the surface of a gravel bar ridge that runs parallel to dunes for ~200 m east-northeast to west-southwest, rising ~1 m above the rest of the plain, and on two mounds ~70 m to the northeast (Fig. 1, Extended Data Fig. 1b, Extended Data Fig. 2 and Supplementary Information 1). Small- to medium-sized gravel covers the surface of the ridge and mounds, lying loosely over a layer of lake sediments. The carbonates, nodules, and shells of gastropods and clams suggest the area was once partly covered by the small lagoon. Most of the fragmentary animal remains recovered are aquatic/lake-edge animals (Supplementary Information 4). The site of Nataruk has relatively few archaeological remains. The excavation of 12 skeletons *in situ* yielded only 131 lithics; however, an area of approximately 6 m × 6 m at the top of the ridge, overlying some of the skeletal remains, had a localized concentration of 628 lithics, and may reflect an ephemeral later occupation of the site (Supplementary Information 5). The lithic industry is similar to other Later Stone Age assemblages in the area^{21–23}, including fragments of barbed bone harpoons typical of early Holocene hunter-fishers of Turkana^{24,25}.

The human skeletal remains had no collagen. Two radiocarbon dates were obtained from sediment samples collected from sub-surface deposits above the skeletons, four from shells found next to or in direct association with the human remains, and another one from within an optically stimulated luminescence (OSL) dating sample collected from the lake sediments in which one of the skeletons was found. As expected, the surface sediments above the skeletons are younger (7,270–8,160 calibrated years before the present (cal. BP)) than the lagoon's shells (9,030–11,750 cal. BP; Extended Data Table 1 and Extended Data Fig. 3a), and correspond to a later period of high lake-levels in the area²⁶. An OSL age of 9,680 ± 805 years obtained from the lake sediments adjacent to one of the skeletons is similar to those of the lagoon shells. Lastly, the majority of 15 uranium-series minimum dates obtained from the skeletons also fall into an age range of early to mid-Holocene, although some results extend back to about 40,000 years ago. Some of these older ages may have been affected by uranium leaching (Extended Data Fig. 3b, c). Although a mid-Upper Pleistocene age for the remains cannot be fully excluded,

¹Leverhulme Centre for Human Evolutionary Studies, Department of Archaeology and Anthropology, University of Cambridge, Fitzwilliam Street, Cambridge CB2 1QH, UK. ²Turkana Basin Institute, Nairobi, Kenya. ³National Museums of Kenya, Jomo Kenyatta House, PO Box 152-30500, Lodwar, Kenya. ⁴Departamento de Prehistoria y Arqueología, UNED, c/ Paseo Senda del Rey, 7, 28040 Madrid, Spain. ⁵National Museums of Kenya, PO Box 40658-00100, Nairobi, Kenya. ⁶Jomo Kenyatta University of Agriculture and Technology, PO Box 62000-00200, Nairobi, Kenya. ⁷Research School of Earth Sciences, The Australian National University, Building 142, Mills Road, Acton, Australian Capital Territory 2601, Australia. ⁸Research Laboratory for Archaeology and the History of Art, University of Oxford, Dyson Perrins Building, South Parks Road, Oxford OX1 3QY, UK. ⁹Research Centre for Human Evolution, Environmental Futures Research Institute, Griffith University, 170 Kessels Road, Nathan, Queensland 4111, Australia. ¹⁰Department of Geology, Anna University, Chennai, Tamil Nadu 600025, India.

*These authors contributed equally to this work.

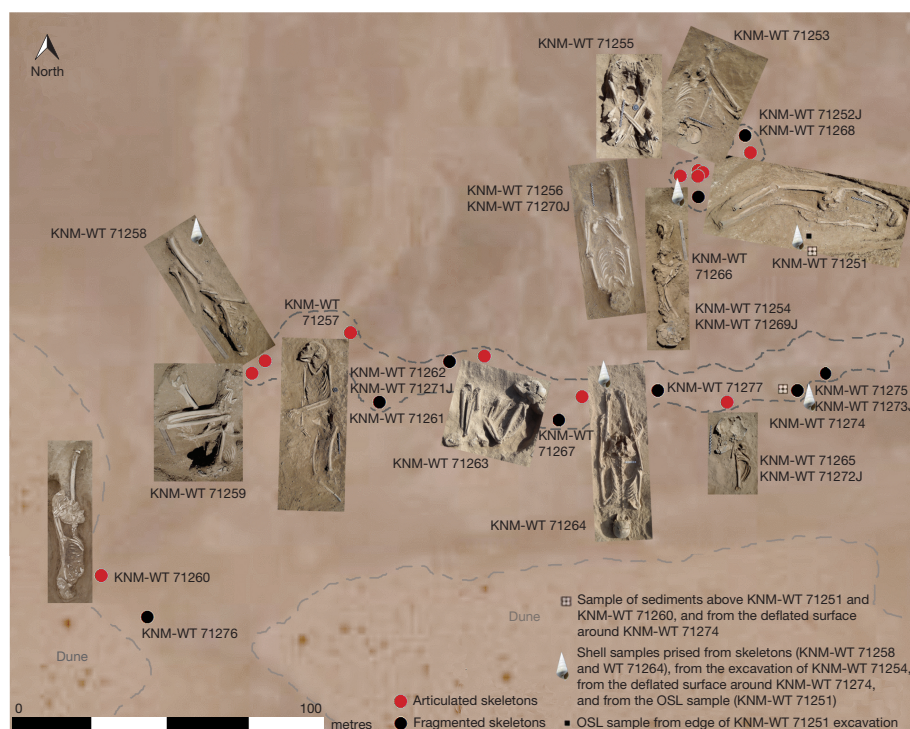


Figure 1 | Spatial distribution of the human skeletons discovered at the site of Nataruk.

Distribution of the remains of 27 individuals found along a sandy ridge and on a small mound at Nataruk; 15 exposed and fragmentary (black), and 12 articulated skeletons (red) illustrated with inset photographs of the position and direction of the bodies as were found. Background image: ArcGIS Online Basemap - World Imagery (source: Esri, DigitalGlobe, GeoEye, i-cubed, USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community).

an age estimate of ~9,500–10,500 years BP for the people of Nataruk is consistent with dates on shells, harpoons, and charcoal from sites in the immediate vicinity, and corresponds to a phase of early Holocene high lake-levels in Turkana³ (Extended Data Table 2 and Supplementary Information 6).

Most of the remains of the people who died at Nataruk were found fully exposed and fragmented, surviving in varying states of preservation and erosion; 12 individuals were partly preserved articulated *in situ*. Among these, no burial pit was identified, and no standardized orientation or position of head, face, or body was observed (Fig. 2). The total number of individuals who died at the site is unknown, as only those partly exposed were excavated. Of the 27 individuals recorded, 21 were adults (8 males, 8 females, and 5 unknown) (Supplementary

Information 2). Partial remains of six children were found commingled or in close proximity to the remains of four adult women and of two fragmentary adults of unknown sex. The remains of a 6- to 9-month-old fetus were recovered from within the abdominal cavity of one of the adult females, representing a 28th individual. No children were found with or near any of the men. All except one of the juvenile remains were children under the age of 6; the exception was a teenager, aged 12–15 years dentally, but whose bones were noticeably small for their age.

Ten of the 12 skeletons *in situ* show evidence of major traumatic lesions that would have been lethal in the immediate- to short-term (Table 1, Extended Data Figs 4–7, Supplementary Information 3 and Supplementary Fig. 1). These include five, possibly six, cases of

Table 1 | Distribution of observable trauma among the articulated human skeletons from Nataruk with more than 20% of the post-cranial skeleton preserved

Individual	Sex	Age	Trauma and other relevant features	Position of lesion(s) and other features
WT 71251	Male	Adult	• Projectile embedded in cranium • Perforating lesion on right parietal bone • Possible knee depressed fractures	• Head • Head • Knee
WT 71253	Male	Adult	• Blunt force trauma on left temporal bone • Perforating lesions on vertebrae	• Head • Neck (×2)
WT 71254	Female?	Adult	• Perforating lesion on frontal bone • Sharp force trauma on mandible • Hand fractures	• Head • Head • Hand
WT 71255	Female	Adult	• Bound? • Pregnant or recently delivered (fetus/newborn)	• Hands
WT 71256	Female	Adult	• Perforating lesion on vertebrae • Hand fracture	• Neck • Hand
WT 71257	Male	Adult	• Blunt force trauma on left temporal bone	• Head
WT 71258	Male	Adult	• Projectiles within body cavity • Bound?	• Thorax • Hands
WT 71259	Female	Adult	• Consecutive rib fractures • Possible knee depressed fractures • Unnatural position of left foot • Bound?	• Ribs • Knee • Foot • Hands
WT 71260	Male	Adult	• Bound?	• Hands
WT 71263	Male	Adult	• Sharp force trauma on frontal bone	• Head
WT 71264	Male	Adult	• Blunt force trauma on left temporal bone	• Head
WT 71265	Female	Adult	• Blunt force trauma on frontal bone	• Head

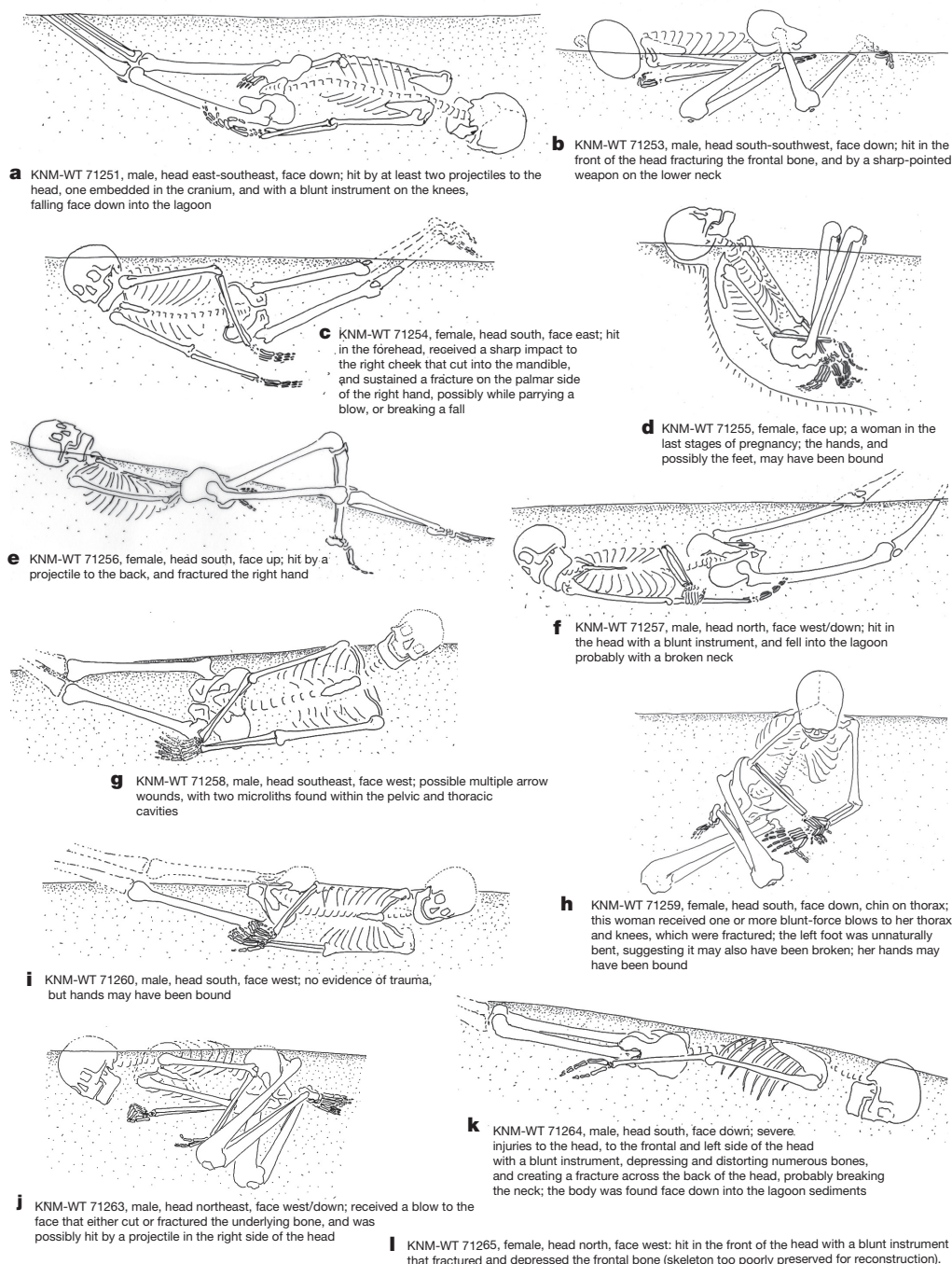


Figure 2 | Schematic drawings illustrating the position of the 12 articulated skeletons from Nataruk. Details of lesions are shown in Extended Data Figs 4–7 and Supplementary Fig. 1, and described in Supplementary Information 3.3.

sharp-force trauma to the head and/or neck probably associated with arrow wounds, five cases of blunt-force trauma to the head, two cases of possible ante-mortem depressed bilateral fractures of the knees, two cases of multiple fractures to the right hand, and a case of fractured ribs. Only two of the skeletons *in situ* show no apparent evidence of peri-mortem trauma, although in both cases the position of the hands suggests the individuals may have been bound at the time of death. In all of the cases of cranial trauma, the compression of bones is localized and cannot be explained by taphonomic forces, as unaffected cranial elements retain the original size and shape around the fractured portions. Three artefacts were found within or embedded in two of the bodies (Fig. 3). The first of these was an obsidian bladelet found embedded in one of the male crania; the others were two microliths, a chert lunate, and an obsidian trapeze, found inside the pelvic and thoracic cavities of a male skeleton; all three showed impact scars.

Interestingly, an obsidian lithic was also found embedded in the foot bones of a skeleton at the nearby site of Lothagam²⁷. The fact that obsidian is relatively rare in other early Holocene Later Stone Age sites of southwest Turkana may suggest that the two groups confronted at Nataruk had different home ranges. The presence of projectile points embedded in the skeletal remains or within the body cavity is considered diagnostic of inter-group conflict, while fractures resulting from blunt and sharp force trauma, particularly to the head, neck, ribs, and hands, are indicative of deliberate violent trauma^{28,29}. A third diagnostic feature of warfare in the past, the presence of cut-marks associated with dismembering and trophy-taking (such as scalping)²⁹, was not observed at Nataruk.

As one of the clearest cases of inter-group violence among prehistoric hunter-gatherers, the event recorded at Nataruk offers information on the socio-economic conditions that marked the presence of warfare.

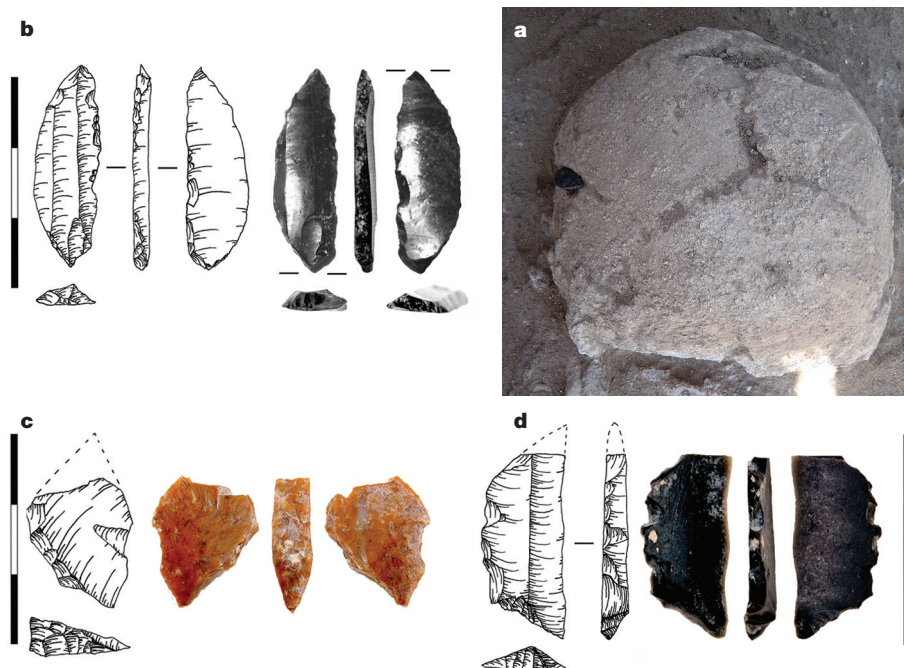


Figure 3 | Lithics found in direct association with human remains at Nataruk. a, KNM-WT 71251 cranium as found *in situ*, with obsidian bladelet found embedded in the left parietal bone. **b,** Detail of obsidian bladelet, showing impact scar at the tip. **c, d,** Microliths found within the

body of KNM-WT 71258, an obsidian crescent, with broken tip, found inside the pelvic basin, and a chert microlith, also with a broken tip, found inside the thoracic cavity. Scale bar units, 1 cm.

However, there are two interpretations of how this fact impinges on our understanding of war among foraging societies. West Turkana 10,000 years ago was a fertile lakeshore landscape sustaining a substantial population of hunter-gatherers; the presence of pottery may be indicative of some storage and so reduced mobility. Thus, the massacre at Nataruk could be seen as resulting from a raid for resources—territory, women, children, food stored in pots—whose value was similar to those of later food-producing societies among whom violent attacks on settlements and organized defence strategies became part of life. In this light, the importance of what happened at Nataruk would be in terms of extending the chronology and degree of the same underlying socio-economic conditions that characterize early warfare in more recent periods. Alternatively, Nataruk may offer evidence not of changing conditions towards a settled, materially richer, and demographically denser way of life, but of a standard antagonistic response to an encounter between two social groups. As such, Nataruk would be important for the particular circumstances that preserved an ephemeral, but perhaps not unusual, event in the life of prehistoric foraging societies. In either case, the deaths at Nataruk are testimony to the antiquity of inter-group violence and war.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 31 July; accepted 23 November 2015.

- Wrangham, R. W. & Glowacki, L. Intergroup aggression in chimpanzees and war in nomadic hunter-gatherers: evaluating the chimpanzee model. *Hum. Nat.* **23**, 5–29 (2012).
- Fry, D. P. & Söderberg, P. Lethal aggression in mobile forager bands and implications for the origins of war. *Science* **341**, 270–273 (2013).
- Garcin, Y. *et al.* Late Pleistocene–Holocene rise and collapse of Lake Suguta, northern Kenya Rift. *Quat. Sci. Rev.* **28**, 911–925 (2009).
- Wilson, M. L. & Wrangham, R. W. Intergroup relations in chimpanzees. *Annu. Rev. Anthropol.* **32**, 363–392 (2003).
- Wilson, M. L. *et al.* Lethal aggression in *Pan* is better explained by adaptive strategies than human impacts. *Nature* **513**, 414–417 (2014).
- Bowles, S. Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? *Science* **324**, 1293–1298 (2009).

- Kelly, R. C. The evolution of lethal intergroup violence. *Proc. Natl Acad. Sci. USA* **102**, 15294–15298 (2005).
- Thorpe, I. J. N. in *Warfare, Violence and Slavery in Prehistory* (eds Parker Pearson, M. & Thorpe, I. J. N.) 1–18 (British Archaeological Reports, 2005).
- Burch, E. S. *Alliance and Conflict: The World System of the Inupiaq Eskimos* (Univ. Nebraska Press, 2005).
- Ember, C. R. Myths about hunter-gatherers. *Ethnology* **17**, 439–448 (1978).
- Fry, D. P. *Beyond War: The Human Potential for Peace*. (Westview, 2007).
- Goldschmidt, W. in *The Social Dynamics of Peace and Conflict* (eds Rubenstein, R. A. & Foster, M. L.) 47–65 (Westview, 1988).
- Hobhouse, L. T., Wheeler, G. C. & Ginsberg, M. *The Material Culture and Social Institutions of the Simpler People* (Chapman & Hall, 1915).
- Leavitt, G. C. The frequency of warfare: an evolutionary perspective. *Sociol. Inq.* **47**, 49–58 (1977).
- Otterbein, K. F. *How War Began* (Texas A&M Univ. Press, 2004).
- Radcliffe-Brown, A. R. *The Andaman Islanders: A Study in Social Anthropology* (Cambridge Univ. Press, 1922).
- Schapera, I. *The Khoisan peoples of South Africa* (Routledge and Kegan Paul, 1930).
- Wright, Q. *A Study of War* (Univ. Chicago Press, 1942).
- Keely, L. H. *War before Civilization* (Oxford Univ. Press, 1996).
- Wendorf, F. in *The Prehistory of Nubia* (ed. Wendorf, F.) Vol. 2, 954–1040 (Southern Methodist Univ. Press, 1968).
- Robbins, L. H. *The Lothagam Site* (Michigan State Univ., 1974).
- Robbins, L. H. Lake Turkana archaeology: the Holocene. *Ethnohistory* **53**, 71–93 (2006).
- Beyin, A. Recent archaeological survey and excavation around the Greater Kalakol area, West Side of Lake Turkana: Preliminary Findings. *Nyame Akuma* **75**, 40–50 (2011).
- Robbins, L. H. Bone artefacts from the Lake Rudolf Basin, East Africa. *Curr. Anthropol.* **16**, 632–633 (1975).
- Barthelme, J. W. *Fisher-Hunters and Neolithic Pastoralists in East Turkana, Kenya* (British Archaeological Reports, 1985).
- Forman, S. L., Wright, D. K. & Blois, C. Variations in water level for Lake Turkana in the past 8500 years near Mt. Por, Kenya and the transition from the African Humid Period to Holocene aridity. *Quat. Sci. Rev.* **97**, 84–101 (2014).
- Angel, J. L., Phenice, T. W., Robbins, L. H. & Lynch, M. M. *Late Stone Age Fishermen of Lothagam, Kenya* (Michigan State Univ., 1980).
- Sauer, N. J. in *Forensic Osteology* (ed. Reichs, K. J.) 321–332 (Charles C. Thomas, 1998).
- Smith, M. O. Beyond palisades: the nature and frequency of late prehistoric deliberate violent trauma in the Chickamauga reservoir of east Tennessee. *Am. J. Phys. Anthropol.* **121**, 303–318 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Office of the President of Kenya, the Turkana Province government, and the National Museums of Kenya for permission to conduct research (NCST/5/002/R/419), the Turkana people of Locher Akwan, Lokwar Ankhaeso, Lotukumo, Eporon, and Natome for permission to work in their area and assistance, the Turkana Basin Institute for logistical and laboratory support, the British Institute in Eastern Africa, R. Leakey, M. Leakey, and L. Martin for support and advice, the staff at the TBI Turkwell facility, especially K. Onesmus Ngela, and the 2012 IN-AFRICA field team (E. Murungi, J. Oltimbao, J. Lokuruka, D. Lomuria, M. Lokinei, J. Ekeno, J. Erupe, J. Lopua, R. Ng'iroton, P. Amuk, P. Atadeit, M. Emusugut, F. Lowan, R. Ng'ichila, S. Eperon, P. Eperon, T. Echulum), especially P. Ebeya who found the site. We thank S. Black and C. Cunningham for advice on the foetal remains, Beta Analytic for advice and assistance, and F. Lahr for assistance with imaging and illustrations. Funding was provided by a European Research Council Advanced Award to M.M.L. (IN-AFRICA, ERC 295907), the Newby Trust, and the McDonald Institute for Archaeological Research, University of Cambridge.

Author Contributions M.M.L. directed the study and fieldwork; M.M.L., F.R., A.M., A.W., J.E.E., J.L., H.M., D.M.M., A.Mu., B.C., H.A., and R.A.F. participated in the fieldwork and excavations; M.M.L., F.R., J.E.E., J.L., H.M., and D.M.M.

excavated the Nataruk skeletons; M.M.L., F.R., R.K.P., A.W., A.L., and K.C. cleaned, prepared, and reconstructed the fossils at the Turkana Basin Institute research laboratories; M.M.L. and R.K.P. analysed and described the lesions and pathologies; M.M.L. prepared the illustrations of Fig. 2 and illustration and photographs of Extended Data Figs 4–7; A.W. prepared the archaeological illustrations and samples for radiocarbon dating; A.M. prepared the three-dimensional model of KNM-WT 71264 in Supplementary Figure 1; M.M.L., A.W., F.R., and F.C. performed the fauna identification; A.W. and J.M.M.F. analysed the lithics following protocols developed by R.A.F., and A.W. wrote the archaeology section of the Supplementary Information; A.V.B. prepared the GIS maps; H.A., R.A.F., M.M.L., and A.M. performed the geomorphological study; R.W. did radiocarbon dating; J.-L.S. did the OSL date; R.G. examined all materials at the Turkana Basin Institute for dating, and did the uranium-series dates; M.M.L. wrote the paper, with contributions from R.A.F., R.G., R.K.P., A.W., E.M., R.W., and J.-L.S.

Author Information The human remains from the site of Nataruk are curated for the National Museums of Kenya (KNM-WT 71251-71277) at the Turkwel Station of the Turkana Basin Institute, Kenya. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.M.L. (mbml1@cam.ac.uk) or R.A.F. (raf10@cam.ac.uk).

METHODS

Core osteological assessment of the skeletal remains from Nataruk. Age at death of juvenile remains was assessed through dental development and epiphyseal fusion^{30–32}. Adult remains were identified cranially by presence of the third molar in occlusion and/or fusion of the sphenoid-occipital synchondrosis, and post-cranially by the fusion of epiphyses and size^{33,34}. Adults were further grouped into young, middle-aged, and old adults depending on degree of dental wear, cranial suture obliteration, presence of degenerative joint disease, and other age-related osteological changes (including pubic symphyses, auricular surfaces, and sternal rib-end morphology)^{35–38}. Sex was determined from pelvic and cranial features following standard practice³⁴. More details in Supplementary Information 2.

Identification of peri-mortem trauma in the skeletal remains from Nataruk. Following standard practice^{28,29,39–43}, the presence of projectile points embedded in the bone or within the body cavity, and peri-mortem fractures resulting from blunt or sharp-force trauma were considered evidence of violence. Antemortem vs peri-mortem trauma was diagnosed on the basis of (1) evidence of healing and/or periosteal reaction^{28,44} and (2) the context of the skeletons and distribution of lesions, noting that violent injuries are most often observed in ribs, scapulae, forearms, hands, and particularly in the head and neck^{45–48}, and that blunt-force peri-mortem traumatic lesions in the head can be considered diagnostic of inter-group conflict in some contexts (for example refs 29, 49–51). Peri-mortem fractures were identified by (1) presence of depressed adhering bone fragments, (2) secondary linear and/or concentric radiating fractures, (3) inner beveling of the margins of the lesion, (4) irregular, sharp or splintered edges (instead of shattered), (5) flaking of the internal fracture surface, and (6) absence of discoloration (that is, the colour of the borders of the lesion is consistent with the surrounding bone)^{28,49,52–62}. Perforating lesions were identified by (1) a linear, circular, or ellipsoidal section, (2) well-defined edges and either smooth or serrated/splintered margins, which may be polished, (3) a V-, semi-V-, or U-shaped cross-section, (4) the presence of internal beveling and/or parallel striations, and (5) anatomical position consistent with a projectile's trajectory. Perforating lesions caused by non-sharp projectiles, such as a bone point or polished arrow shaft, were identified by typically rounded or oval margins, with rounded edges⁶³. Further information on the traumatic lesions observed in Nataruk in Supplementary Information 3.

Radiocarbon dating. Two sediment samples from above two of the skeletons were dated by Beta Analytic. These were sieved to <180 µm, and the resulting bulk organic fraction acid-washed to remove carbonates in a series of 1.0 N HCl leaches at 90 °C for ≥1.5 h, followed by serial rinses in de-ionized water at 70 °C, drying in an oven at 90 °C for 12–24 h, and further homogenization and HCl sub-sample application to validate the absence of carbonates. Microscopic examination of the entire remaining sample insured removal of any root hairs and/or fragments. The acid-insoluble organic fraction of a sub-sample was submitted for accelerator mass spectrometry dating without separation of the humic/humin fractions; no stepped combustion process was performed. Five radiocarbon dates were obtained from gastropod shells. Four individual shells were dated by Beta Analytic; pre-treatment involved removal of the surface through acid etching (BA-344846 13.4 mg; BA-344849 9.7 mg; BA-3344847 13.0 mg; BA-3344848 11.4 mg). A small (12.4 mg), weathered gastropod was dated at the Australian National University (S-ANU 37218). The protocol included (1) removing the surface with a scalpel and leaching the remaining sample in HCl until 10 wt% was lost to exclude recrystallized material; (2) powdering the sample and submitting it to X-ray diffraction analysis to test for the presence of calcite, which under normal conditions precipitates from solution as calcite and can be distinguished from the aragonite form of carbonate in most gastropods. X-ray diffraction was undertaken in a Siemens D501 diffractometer at the ANU, operating at 40 mA and 40 kV, using Cu K α radiation with a step size and time of 0.5 2 θ and 60 s between 25 and 50 2 θ on the Bragg scale, and Siroquant to quantify the calcite content; (3) reacting the powdered sample with 85% H₃PO₄ in an evacuated Vacutainer, and collecting and purifying the CO₂ generated cryogenically before conversion to graphite with hydrogen over an iron catalyst; (4) performing accelerator mass spectrometry dating following procedures specified in ref. 64 and subtracting sample preparation backgrounds on the basis of measurements of samples of ¹⁴C-free CO₂. All radiocarbon dates were calibrated using OxCal version 4.2.4 (ref. 65), IntCal13 curve^{66,67}. Further information on radiocarbon dating is given in Supplementary Information 6.2.

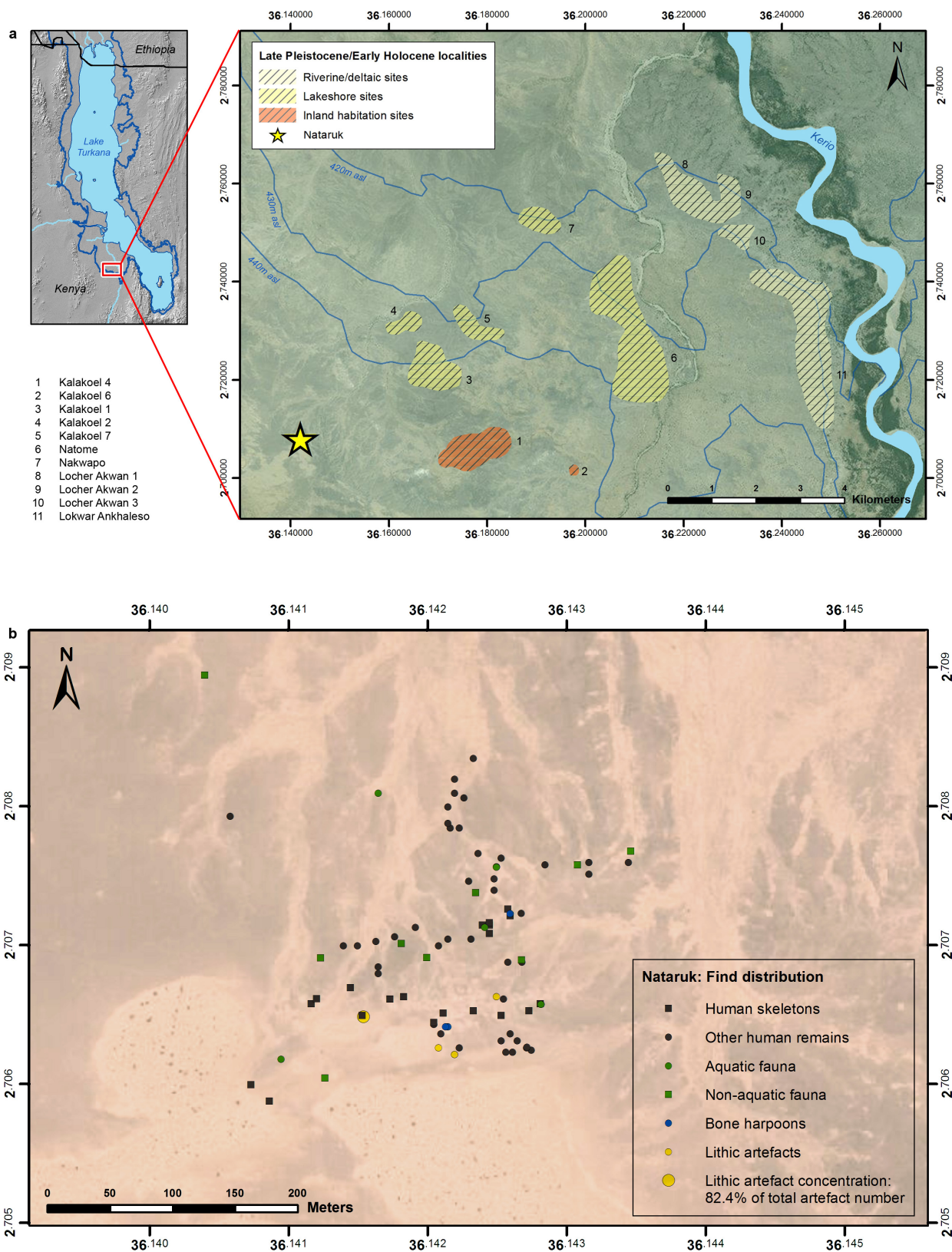
OSL dating. An OSL date was obtained at the University of Oxford (X6481) from a block sample of the semi-consolidated sediment that was exposed through the excavation of KNM-WT 71251. Sample preparation for OSL dating involved the removal of the light-exposed outer 2 cm of the block under filtered laboratory lighting (low intensity LEDs with peak emission at 559 nm) and the extraction of sand-sized (90–125 µm) quartz mineral grains (prepared through wet sieving, HCl acid digestion, etching in HF acid (45%), and heavy liquid flotation (sodium polytungstate)). Owing to the common occurrence of feldspar mineral grains

within the mineral matrix, the sample was also subjected to a prolonged (2 weeks) etching in H₂SiF₆. Purified quartz grains were mounted as multi-grain mono-layers of circa 4 mm diameter onto aluminium discs with a silicone oil adhesive. OSL measurements were conducted using an automated Riso luminescence reader⁶⁷, based on a conventional single-aliquot regeneration measurement protocol^{68,69}. To minimize the contribution of residual feldspathic components to the quartz signal, each OSL measurement was preceded by an infrared bleach^{70,71}. Optical stimulation for single aliquots was provided by clusters of blue light emitting diodes (42 Nichia 470Δ20nm) providing a sample stimulation power of ~32–36 mW cm⁻². The natural and regenerative doses were preheated to 240 °C for 10 s, and the fixed test doses used to correct for sensitivity changes were preheated to a reduced temperature of 220 °C, before optical stimulation. The choice of preheat combination followed a series of dose recovery experiments conducted at 210, 230, 240, and 250 °C. These showed no dependency on temperature, returning equivalent dose values close to unity (0.97, 1.03, 0.96, and 1.04). The presence of infrared-sensitive minerals (for example, feldspars) was checked using an infrared bleach provided by a solid-state laser diode (830Δ10 nm; 1 W cm⁻²) at 50 °C for 50 s before blue light stimulation. The ultraviolet quartz OSL emission at ~370 nm used for optical dating was detected using an Electron Tubes 9235QA photomultiplier tube fitted with a blue-green sensitive bialkali photocathode and a 7.5 mm Hoya U-340 glass filter. Laboratory doses used for constructing the dose–response curves were provided by a 90Sr/90Y ceramic beta source housed within the reader and calibrated against a gamma irradiated Riso National Laboratory standard⁷². The standard error on individual equivalent dose measurements included an instrument reproducibility uncertainty of 1%, as well as a random 1% uncertainty arising from photon counting statistics. The total uncertainty on the final equivalent dose included a further systematic component of 4% (added in quadrature) to account for uncertainties in the calibration of the inbuilt beta source. The equivalent dose was determined from the first second of the OSL decay curve (Extended Data Fig. 3d) using the final 5 s as background noise (total stimulation time was 50 s). Dose–response curves (Extended Data Fig. 3e) were fitted with the Analyst software package⁷³ using a double saturating exponential function; the distribution of replicate equivalent dose measurements is presented in Extended Data Fig. 3f. Given the internal consistency between replicate equivalent dose measurements and the otherwise favourable OSL characteristics of the multigrain quartz aliquots (Supplementary Information 6.3), a central age model⁷⁴ was used to obtain the final equivalent dose estimate for age calculation. Dose rate calculations were based on the concentrations of radioactive elements (potassium, thorium, and uranium) within the sample as determined from elemental analysis performed on homogenized and pulverized sub-samples (approximately 10 g of sediment) by inductively coupled mass spectrometry and inductively coupled atomic emission spectroscopy. No direct evaluation of ²³⁸U and ²³²Th within the quartz grains was made as the internal dose rate only provides a very minor contribution (<300 µGy kyr⁻¹) to the total dose rate^{75–77} and is often considered to be negligible. However, an assumed internal (alpha and beta) dose rate of 0.03 ± 0.01 Gy kyr⁻¹ based on more recently published ²³⁸U and ²³²Th measurements for etched quartz^{78,79} and an alpha efficiency factor of 0.04 ± 0.01 (ref. 80) was included in the dose rate calculations. Owing to technical difficulties in measuring radionuclide concentrations in quartz extracts, a more meticulous and laborious evaluation of the contribution of the internal dose rate was not considered necessary, especially because the OSL dating was not aimed at establishing a high-resolution chronology. Furthermore, because of the relatively high external dose rate (~2 Gy kyr⁻¹) encountered in this particular environmental setting, there was little benefit from making a more precise determination. The uncertainty associated with the internal dose rate was not considered to have a significant effect on the reported age estimate. The concentrations of parent isotopes were converted to dose rates according to the updated attenuation factors proposed in ref. 81, corrections for grain size⁷⁶, as well as water content⁸². The recorded moisture content from the sample was below 1%. However, this value was not considered a good approximation of the mean water content of the sediment throughout the burial period. To account for past and seasonal changes in the pore water content, a mean long-term value of 5% of the dry mass of the sample was used and assigned a relative uncertainty of 60% for determining the attenuating effect on the dose rate received by the quartz mineral grains. The cosmic-ray dose was calculated according to ref. 83, taking into account the thickness and density of the overburden as well as the geomagnetic latitude and elevation of the site. Despite the current erosional nature of the landscape, the recorded shallow overburden thickness of 40 cm is considered to represent the best long-term approximation. There is no evidence to suggest that the site was ever covered by nearby dunes, which are believed to have accumulated relatively recently. Observations on the OSL dating of sample X6481 from Nataruk is given in Supplementary Information 6.3.

Uranium-series dating. Laser ablation uranium-series analyses were performed on 18 skeletal samples following principles and procedures described in ref. 84.

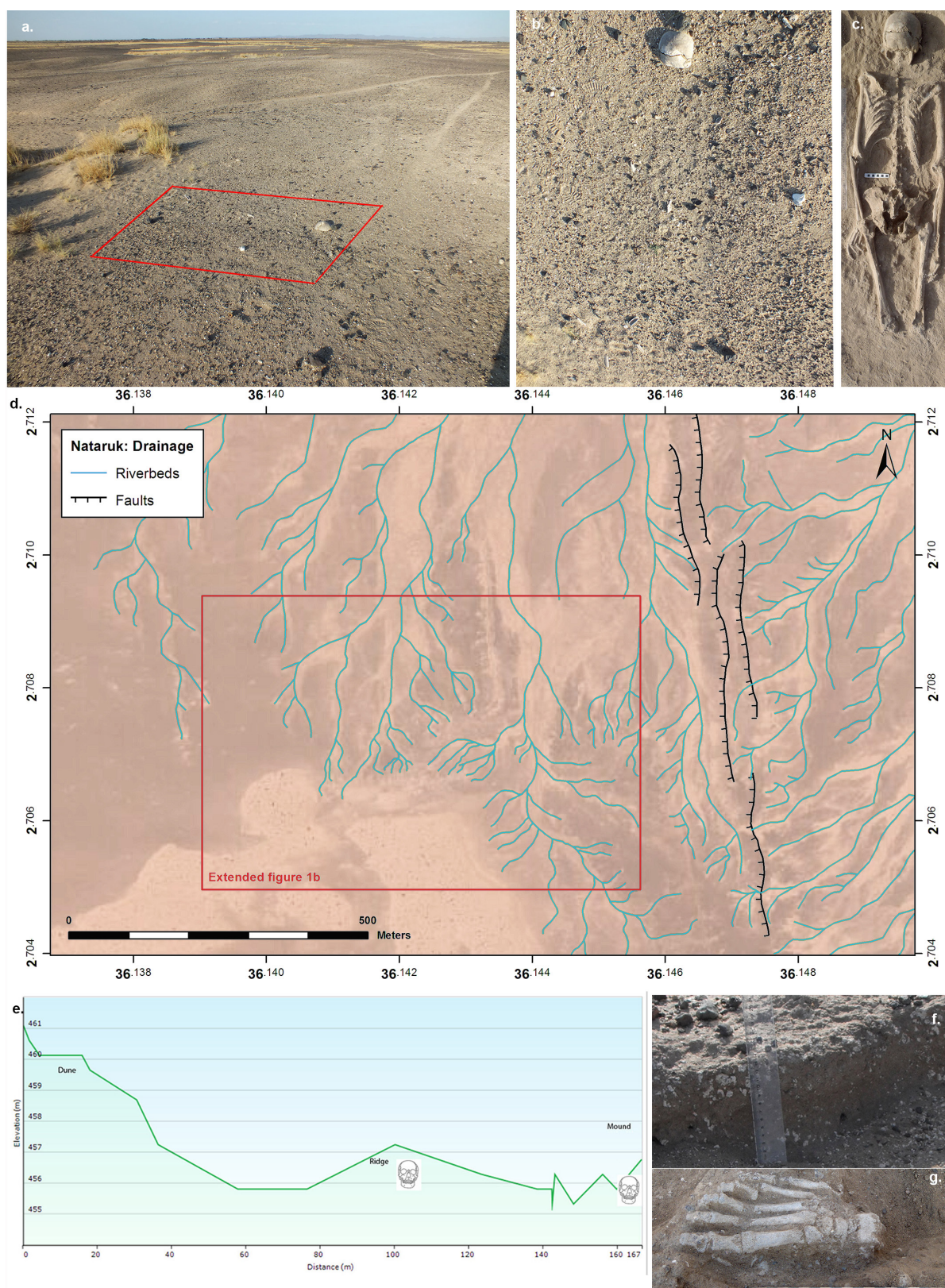
Laser ablation data were obtained from spot analyses collected on the cortical sections of the bone on lines perpendicular to the surface. On thick cortical bones, data were collected along a single line; on samples with thinner cortical bone sections (for example, ribs with a spongy centre), spots were obtained along several lines. No individual age calculation was performed when the uranium concentrations were below about 0.5 p.p.m. and detrital ^{232}Th was observed⁸⁴ (elemental U/Th ratios below 100). If there were more than four remaining data points, these were used for the calculation of diffusion-adsorption-decay age estimates⁸⁵. Diffusion-adsorption-decay calculations were only considered successful if there was no conflict between the $^{230}\text{Th}/^{238}\text{U}$ and $^{234}\text{U}/^{238}\text{U}$ data across the bone. In cases where these conditions were not met, the average spot data were used to calculate average ages. Detailed information on the uranium-series results is given in Supplementary Information 6.4.

30. Broadbent, B. H. & Golden, W. H. *Bolton Standards of Dentofacial Developmental Growth* (CV Mosby, 1975).
31. Smith, B. H. in *Advances in Dental Anthropology* (eds Kelley, M. A. & Larsen, C. S.) 143–168 (Wiley-Liss, 1991).
32. Scheuer, L. & Black, S. M. *Developmental Juvenile Osteology* (Academic, 2000).
33. Bass, W. M. *Human Osteology: A Laboratory and Field Manual* (Missouri Archaeological Society, 1995).
34. Buikstra, J. E. & Ubelaker, D. H. *Standards for Data Collection from Human Skeletal Remains* (Arkansas Archaeological Survey, 1994).
35. Brooks, S. & Suchey, J. M. Skeletal age determination based on the os pubis: a comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. *Hum. Evol.* **5**, 227–238 (1990).
36. Brothwell, D. R. *Digging Up Bones* (Oxford Univ. Press, 1981).
37. İscan, M. Y., Loth, S. R. & Wright, R. K. Metamorphosis at the sternal rib end: a new method to estimate age at death in white males. *Am. J. Phys. Anthropol.* **65**, 147–156 (1984).
38. İscan, M. Y., Loth, S. R. & Wright, R. K. Age estimation from the rib by phase analysis: white females. *J. Forensic Sci.* **30**, 853–863 (1985).
39. Brink, O., Vesterby, A. & Jensen, J. Pattern of injuries due to interpersonal violence. *Injury* **29**, 705–709 (1998).
40. Jurmain, R. et al. Paleoepidemiological patterns of interpersonal aggression in a prehistoric central California population from CA-ALA-329. *Am. J. Phys. Anthropol.* **139**, 462–473 (2009).
41. Kjærulff, H. et al. The Copenhagen Study Group. Injuries due to deliberate violence in areas of Denmark. III. Lesions. *Forensic Sci.* **41**, 169–180 (1989).
42. Martin, D. L. & Frayer, T. *Troubled Times: Osteological and Anthropological Evidence of Violence* (Gordon and Breach, 1997).
43. Shepherd, J. P., Shapland, M., Pearce, N. X. & Scully, C. Pattern, severity and aetiology of injuries in victims of assault. *J. R. Soc. Med.* **83**, 75–78 (1990).
44. Quatrehomme, G. & İscan, M. Y. Postmortem skeletal lesions. *Forensic Sci. Int.* **89**, 155–165 (1997).
45. Angel, J. L. Patterns of fracture from Neolithic to modern times. *Anthropol. Közlemények* **18**, 9–18 (1974).
46. Brickley, M. & Smith, M. Culturally determined patterns of violence: biological anthropological investigations at a historic urban cemetery. *Am. Anthropol.* **108**, 163–177 (2006).
47. Fibiger, L., Ahlström, T., Bennike, P. & Schulting, R. J. Patterns of violence-related skull trauma in Neolithic southern Scandinavia. *Am. J. Phys. Anthropol.* **150**, 190–202 (2013).
48. Levin, L. et al. Incidence and severity of maxillofacial injuries during the Second Lebanon War among Israeli soldiers and civilians. *J. Oral Maxillofac. Surg.* **66**, 1630–1633 (2008).
49. Steadman, D. W. Warfare related trauma at Orendorf, a middle Mississippian site in west-central Illinois. *Am. J. Phys. Anthropol.* **136**, 51–64 (2008).
50. Owsley, D. W., Berryman, H. E. & Bass, W. M. Demographic and osteological evidence for warfare at the Larson site, South Dakota. *Plains Anthropological Memoirs* **13**, 119–131 (1977).
51. Walker, P. L. in *Troubled Times: Violence and Warfare in the Past* (eds Martin, D. & Frayer, D. W.) 145–180 (Gordon and Breach, 1997).
52. Berryman, H. E. & Haun, S. J. Applying forensic techniques to interpret cranial fracture patterns in an archaeological specimen. *Int. J. Osteoarchaeol.* **6**, 2–9 (1996).
53. Berryman, H. E. & Symes, S. A. in *Forensic Osteology* (ed. Reichs, K. J.) 333–352 (Charles C. Thomas, 1998).
54. Galloway, A. in *Broken Bones: Anthropological Analysis of Blunt Force Trauma* (ed. Galloway, A.) 81–112 (Charles C. Thomas, 1999).
55. Kimmerle, E. H. & Baraybar, J. P. *Skeletal Trauma: Identification of Injuries Resulting from Human Rights Abuse and Armed Conflict* (Taylor & Francis, 2008).
56. Knüsel, C. J. in *Warfare, Violence and Slavery in Prehistory* (eds Parker Pearson, M. & Thorpe, I. J. N.) 49–65 (British Archaeological Reports, 2005).
57. Lovell, N. C. Trauma analysis in paleopathology. *Am. J. Phys. Anthropol.* **104**, 139–170 (1997).
58. Ortner, D. in *Skeletal Trauma: Identification of Injuries Resulting from Human Rights Abuse and Armed Conflict* (eds Kimmerle, E. H. & Baraybar, J. P.) 21–86 (CRC, 2008).
59. Passalacqua, N. V. & Fenton, T. W. in *A Companion to Forensic Anthropology* (ed. Dirkmat, D. C.) 400–411 (Blackwell, 2012).
60. Roberts, C. in *Human Paleopathology: Current Synthesis and Future Options* (eds Ortner, D. & Aufderheide, A. C.) 225–240 (Smithsonian Institution Press, 1991).
61. Rodríguez-Martín, C. in *Forensic Anthropology and Medicine* (eds Schmitt, A. et al.) 197–221 (Humana, 2006).
62. Wheatley, B. P. Perimortem or postmortem bone fractures? An experimental study of fracture patterns in deer femora. *J. Forensic Sci.* **53**, 69–72 (2008).
63. Letourneau, C. & Petillon, J. Hunting lesions caused by osseous projectile points: experimental results and archaeological implications. *J. Archaeol. Sci.* **35**, 2849–2862 (2008).
64. Fallon, S. J., Fifield, L. K. & Chappell, J. M. The next chapter in radiocarbon dating at the Australian National University: status report on the single stage accelerator mass spectrometry. *Nucl. Instrum. Methods Phys. Res. B* **268**, 898–901 (2010).
65. Bronk Ramsey, C., Scott, E. M. & van der Plicht, J. Calibration for archaeological and environmental terrestrial samples in the time range 26–50 ka cal BP. *Radiocarbon* **55**, 2021–2027 (2013).
66. Reimer, P. et al. IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).
67. Bøtter-Jensen, L. Luminescence techniques: instrumentation and methods. *Radiat. Meas.* **27**, 749–768 (1997).
68. Murray, A. S. & Wintle, A. G. Luminescence dating of quartz using an improved single-aliquot regenerative-dose protocol. *Radiat. Meas.* **32**, 57–73 (2000).
69. Wintle, A. G. & Murray, A. S. A review of quartz optically stimulated luminescence characteristics and their relevance in single-aliquot regeneration dating protocols. *Radiat. Meas.* **41**, 369–391 (2006).
70. Banerjee, D., Murray, A. S., Bøtter-Jensen, L. & Lang, A. Equivalent dose estimation using a single aliquot of polymineral fine grains. *Radiat. Meas.* **33**, 73–94 (2001).
71. Wallinga, J., Murray, A. S. & Bøtter-Jensen, L. Measurement of the dose in quartz in the presence of feldspar contamination. *Radiat. Prot. Dosim.* **101**, 367–370 (2002).
72. Hansen, V., Murray, A. S., Buylaert, J. P., Yeo, E. Y. & Thomsen, K. J. A new irradiated quartz for beta source calibration. *Radiat. Measur.* **81**, 123–127 (2015).
73. Duller, G. A. T. The Analyst software package for luminescence data: overview and recent improvements. *Anc. TL* **33**, 35–42 (2015).
74. Galbraith, R. F., Roberts, R. G., Laslett, G. M., Yoshida, H. & Olley, J. M. Optical dating of single and multiple grains of quartz from Jinnium Rock Shelter, northern Australia: part I, experimental design and statistical methods. *Archaeometry* **41**, 339–364 (1999).
75. Sutton, S. R. & Zimmermann, D. W. Thermoluminescence dating: radioactivity in quartz. *Archaeometry* **20**, 67–69 (1978).
76. Mejdahl, V. Thermoluminescence dating: beta-dose attenuation in quartz grains. *Archaeometry* **21**, 61–72 (1979).
77. Grün, R. & Fenton, C. Internal dose rates of quartz grains separated from fault gouge. *Anc. TL* **29**, 26–28 (1990).
78. De Corte, F., Vandenberghe, D., Buylaert, J. P., Van den Haute, P. & Kucera, J. Relative and k0-standardized INAA to assess the internal (Th, U) radiation dose rate in the 'quartz coarse-grain protocol' for OSL dating of sediments: Unexpected observations. *Nucl. Instrum. Methods Phys. Res. A* **564**, 743–751 (2006).
79. Vandenberghe, D., De Corte, F., Buylaert, J. P., Kucera, J. & Van den Haute, P. On the internal radioactivity in quartz. *Radiat. Meas.* **43**, 771–775 (2008).
80. Rees-Jones, J. & Tite, M. S. Optical dating results for British archaeological sediments. *Archaeometry* **39**, 177–187 (1997).
81. Guerin, G., Mercier, N. & Adamiec, G. Dose-rate conversion factors: update. *Anc. TL* **29**, 5–8 (2011).
82. Zimmerman, D. W. Thermoluminescent dating using fine grains from pottery. *Archaeometry* **13**, 29–52 (1971).
83. Prescott, J. R. & Hutton, J. T. Cosmic ray contributions to dose rates for luminescence and ESR dating: large depths and long-term time variations. *Radiat. Meas.* **23**, 497–500 (1994).
84. Grün, R., Eggins, S., Kinsley, L., Mosely, H. & Sambridge, M. Laser ablation U-series analysis of fossil bones and teeth. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **416**, 150–167 (2014).
85. Sambridge, M., Grün, R. & Eggins, S. U-series dating of bone in an open system: the diffusion-adsorption-decay model. *Quat. Geochronol.* **9**, 42–53 (2012).
86. Dietze, M. et al. The abanico plot: visualizing chronometric data with individual standard errors. *Quat. Geochronol.* (in the press).
87. Kreutzer, S. et al. Introducing an R package for luminescence dating analysis. *Anc. TL* **30**, 1–8 (2012).
88. R Development Core Team R. A Language and Environment for Statistical Computing. <http://CRAN.R-project.org>.



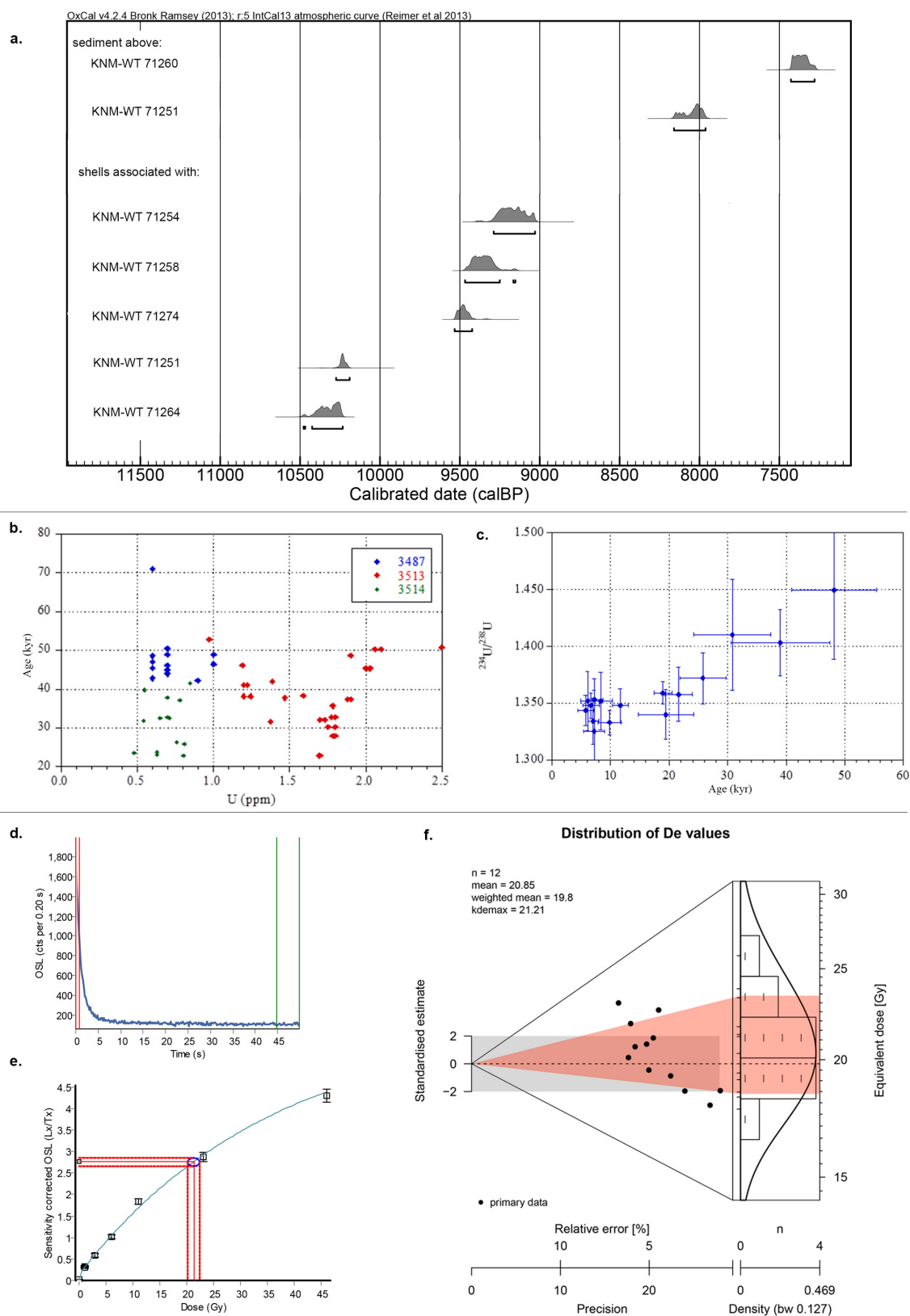
Extended Data Figure 1 | Location of Nataruk and distribution of finds within the site. **a**, Geographical relation of the site of Nataruk to other similarly dated sites in the area and to the reconstructed palaeo-shorelines of Lake Turkana at different extents during the early Holocene transgression (440 m above sea level (a.s.l.), 430 m a.s.l., 420 m a.s.l.), on the basis of SRTM 90 m DEM (version 4.1) (source: CGIAR-CSI, <http://srtm.csi.cgiar.org>); background image: hillshade based on SRTM

90 m DEM (version 4.1) (source: CGIAR-CSI, <http://srtm.csi.cgiar.org>). **b**, Distribution of finds (human, animal and archaeological remains) at Nataruk; background image: ArcGIS Online Basemap - World Imagery (version 2 October, 2014) (source: Esri, DigitalGlobe, GeoEye, i-cubed, USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community).



Extended Data Figure 2 | The site of Nataruk. **a, b,** View of the Nataruk site with exposed cranium of KNM-WT 71264 as was first found. **c,** KNM-WT 71264 after excavation. **d,** Geomorphological setting of the site of Nataruk, showing pattern of drainage and the small geological fault to the east; background image: ArcGIS Online Basemap - World Imagery (version 2 October, 2014) (source: Esri, DigitalGlobe, GeoEye, i-cubed,

USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community). **e,** Elevation profile of the site of Nataruk showing the position of the ridge and northeast mounds. **f,** View of the lagoon carbonate nodules and shells. **g,** Example of human remains in the context of the lagoon sediments, illustrated by the left foot of KNM-WT 71255. Photographs by M.M.L.

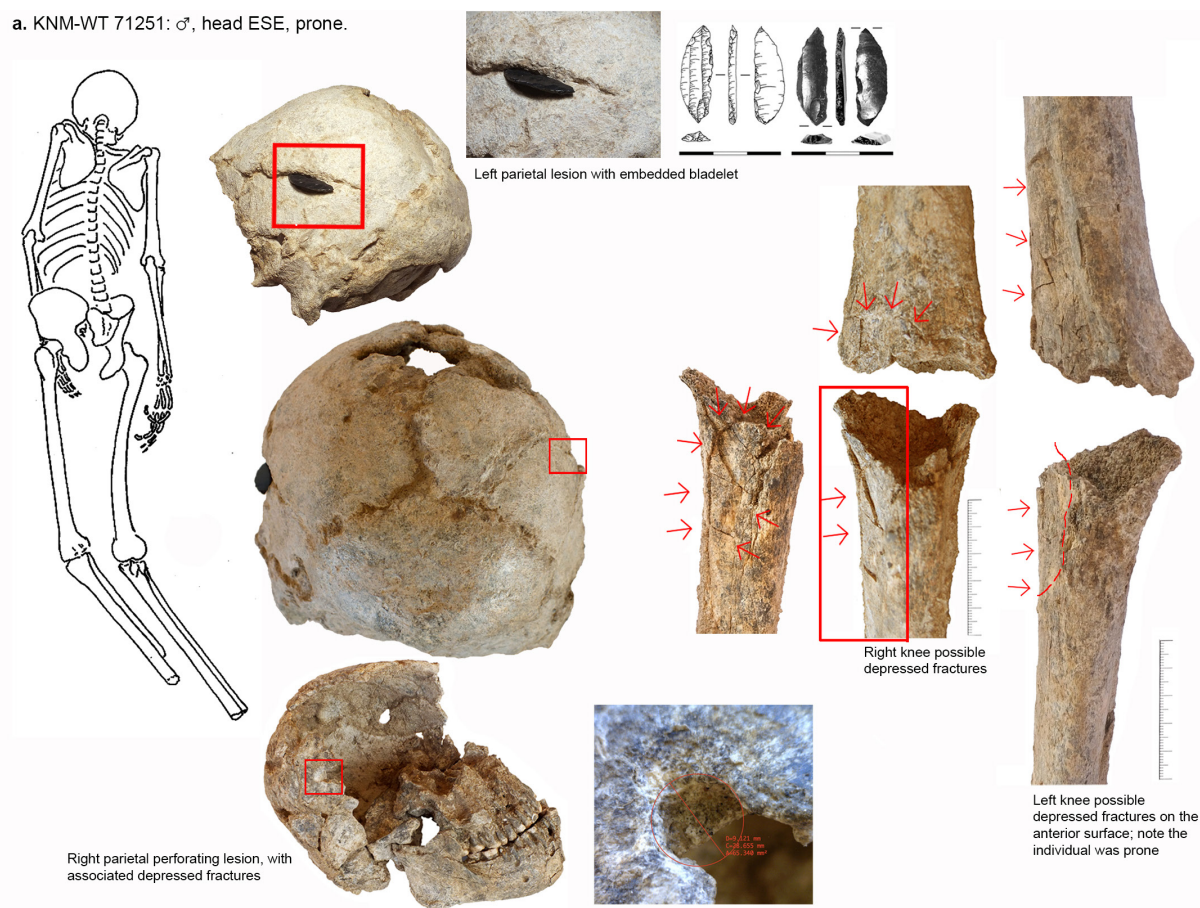


Extended Data Figure 3 | See next page for figure caption.

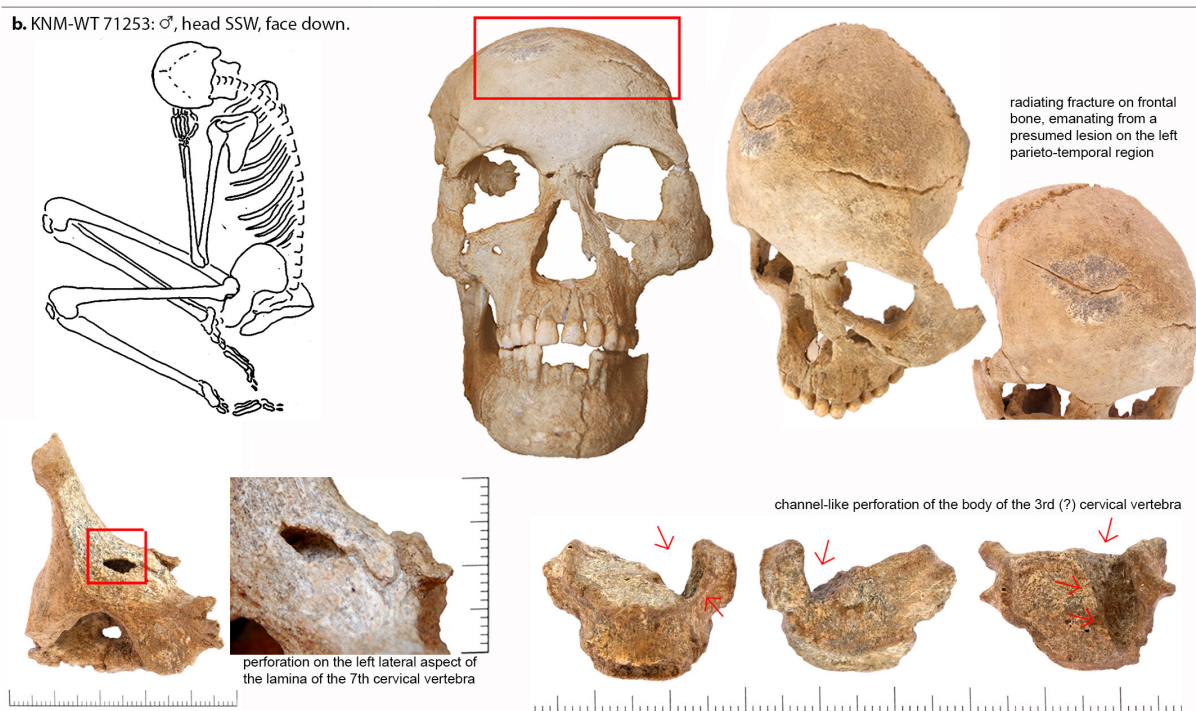
Extended Data Figure 3 | Dating of the site of Nataruk and graphic synthesis of OSL methodology used. **a**, Modelled date (BP) for the sediments above skeletons KNM-WT 71251 and 71260, and the shells associated with skeletons KNM-WT 71251, 71254, 71258, 71264, 71274 showing the age and probability distributions. Sample source and association with human remains to the left of individual plots, with the two younger sediment samples from the surface above skeletons at the top, and the shell samples in direct association with human remains below. **b**, Distribution of estimated uranium-series ages \times uranium (p.p.m.) for samples 3487 and 3513 (both from skeleton KNM-WT 71264) and 3514 (KNM-WT 71265). **c**, Average apparent uranium-series age (error bars indicate 1σ standard deviation) and $^{234}\text{U}/^{238}\text{U}$ ratios of all samples analysed ($n = 17$) (further information in Extended Data Table 3

and Supplementary Table 8). **d**, Example of a natural OSL decay curve featuring the integrated signal (in red) and the background counts (in green) obtained from a multigrain quartz aliquot. **e**, Example of a sensitivity corrected dose–response curve for the natural (in red) and regenerative-dose signals using a double exponential fitting procedure. **f**, Abanico plot featuring the distribution of multigrain quartz equivalent dose measurements and their associated data precision and error scatter⁸⁶. The plot combines a radial plot (bivariate plot on the left side) with a histogram and kernel density estimate curve (univariate plots on the right side) using the default function tool developed within the package ‘Luminescence’⁸⁷ for the statistical programming language ‘R’⁸⁸. The 2σ dispersion range is shown in grey and the red polygon characterizes the 1σ frequency distribution of the primary data.

a. KNM-WT 71251: ♂, head ESE, prone.

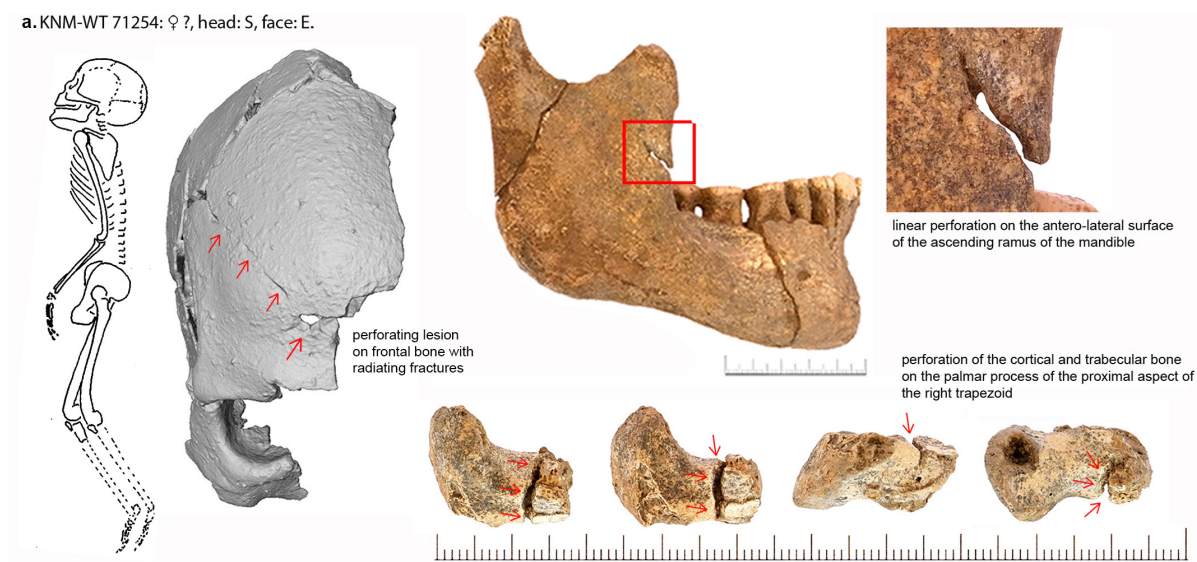
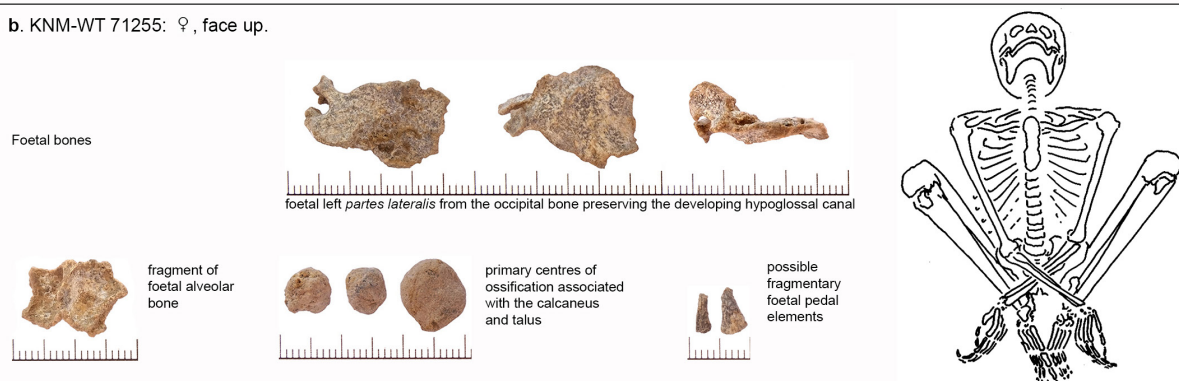
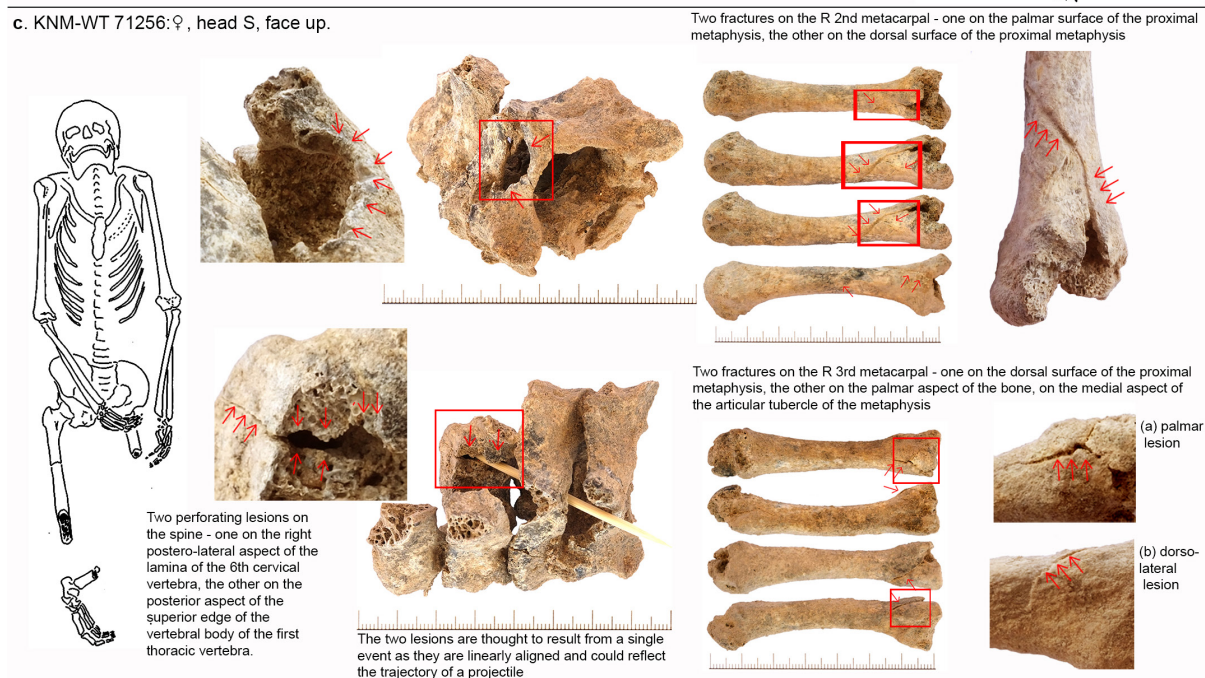


b. KNM-WT 71253: ♂, head SSW, face down.



Extended Data Figure 4 | Traumatic lesions on skeletons. a, Skeleton KNM-WT 71251; **b,** skeleton KNM-WT 71253. Displayed lesions on KNM-WT 71251 include an embedded projectile on the left parietal bone, a perforating lesion on the right parietal bone, and possible depressed fractures on the left and right knees. Displayed lesions on KNM-WT 71253 include radiating fractures on the frontal bone, and perforating

lesions on the third (?) and seventh cervical vertebrae. Red boxes indicate enlarged images of particular lesions; red arrows point to the lesions described in Supplementary Information 3.3.1 and 3.3.2; red dotted line in KNM-WT 71251 delineates the depressed region on the anterior aspect of the left proximal tibia. Photographs and drawings by M.M.L.

a. KNM-WT 71254: ♀?, head: S, face: E.**b. KNM-WT 71255: ♀, face up.****c. KNM-WT 71256: ♀, head S, face up.**

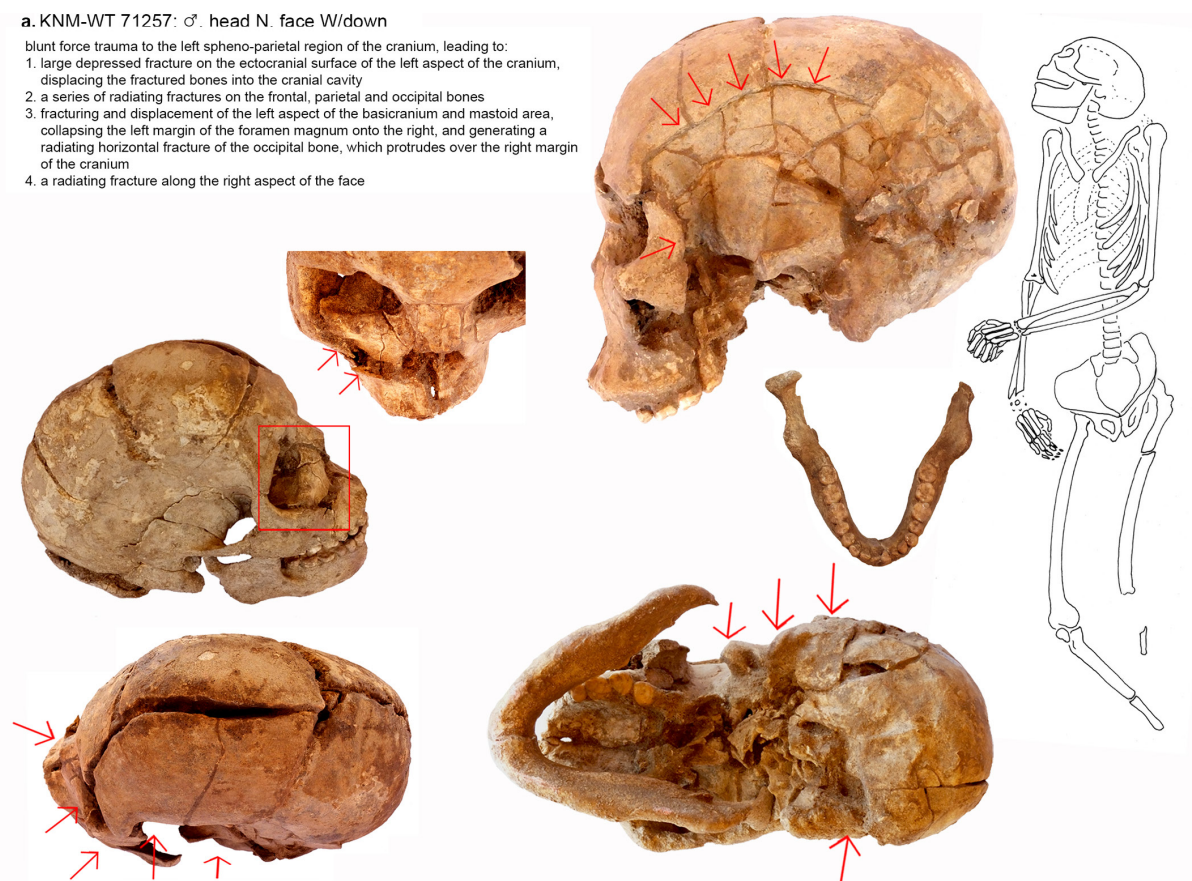
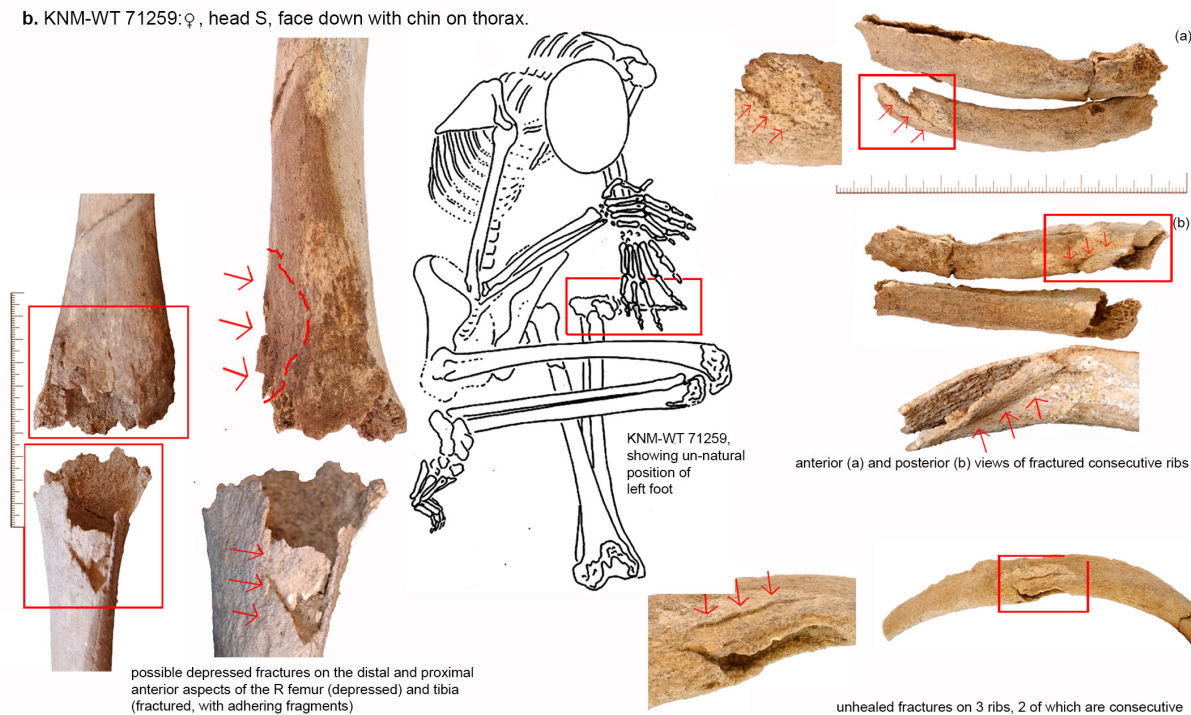
Extended Data Figure 5 | Traumatic lesions on skeletons. a, Skeleton KNM-WT 71254; **c,** skeleton KNM-WT 71256; **b,** fetal remains found within the body cavity of KNM-WT 71255. Displayed lesions on KNM-WT 71254 include a perforating lesion on the frontal bone, and linear perforations on the mandible and the right trapezoid. Displayed lesions on KNM-WT 71256 include perforating lesions on cervical and thoracic vertebrae, and fractures on the second and third right metacarpals. Metacarpal images show four views of the element to

illustrate the extent of the fractures; inset of second metacarpal shows palmar view of lesion, insets of third metacarpal show palmar (a) and dorso-lateral (b) fractures. Red boxes indicate enlarged images of particular lesions; red arrows point to the lesions described in Supplementary Information 3.3.3 and 3.3.5. Further information on KNM-WT 71255 and fetal remains is given in Supplementary Information 3.3.4. Photographs and drawings by M.M.L.

a. KNM-WT 71257: ♂, head N, face W/down

blunt force trauma to the left sphenoparietal region of the cranium, leading to:

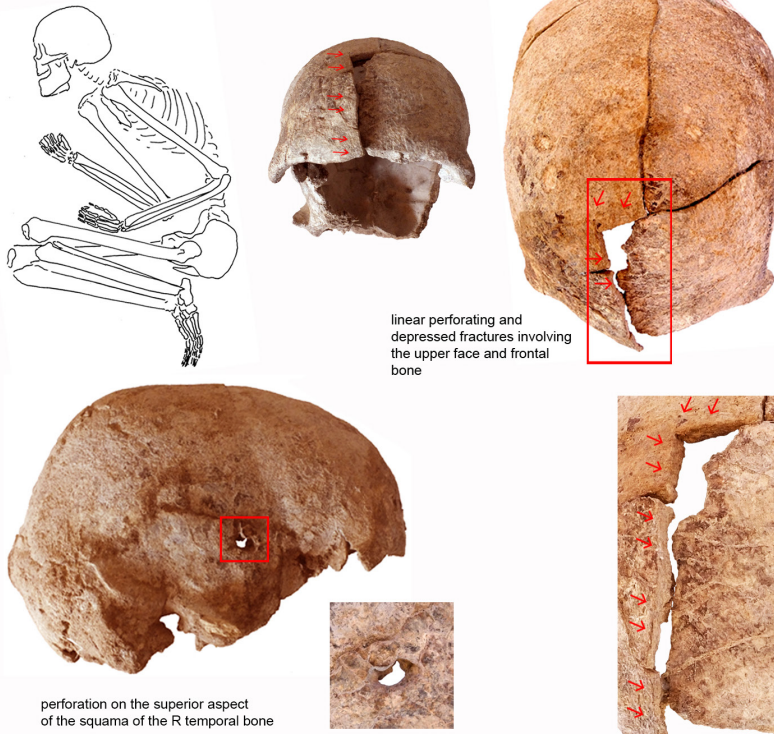
1. large depressed fracture on the ectocranial surface of the left aspect of the cranium, displacing the fractured bones into the cranial cavity
2. a series of radiating fractures on the frontal, parietal and occipital bones
3. fracturing and displacement of the left aspect of the basicranium and mastoid area, collapsing the left margin of the foramen magnum onto the right, and generating a radiating horizontal fracture of the occipital bone, which protrudes over the right margin of the cranium
4. a radiating fracture along the right aspect of the face

**b. KNM-WT 71259: ♀, head S, face down with chin on thorax.**

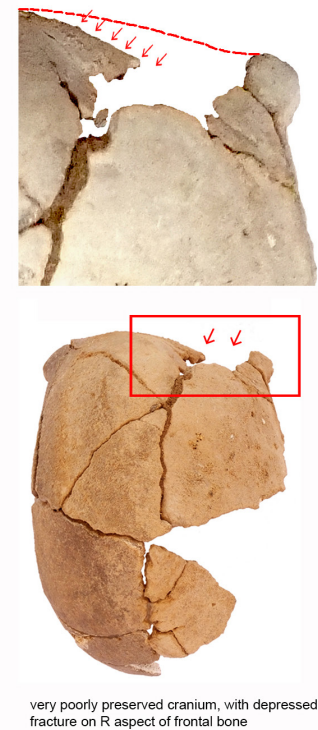
Extended Data Figure 6 | Traumatic lesions on skeletons. a, Skeleton KNM-WT 71257; **b,** skeleton KNM-WT 71259. Displayed lesions on KNM-WT 71257 include depressed fractures on the left parietal and multiple radiating fractures across the skull, with the mandible unaffected. Displayed lesions on KNM-WT 71259 include three fractures to ribs, including on two consecutive ribs, and possible depressed fractures to

the right knee. Red boxes indicate enlarged images of particular lesions in KNM-WT 71259, including the unnatural position of the left foot; red arrows point to the lesions described in Supplementary Information 3.3.6 and 3.3.8; red dotted line in KNM-WT 71259 delineates the depressed region on the anterior aspect of the right distal femur. Photographs and drawings by M.M.L.

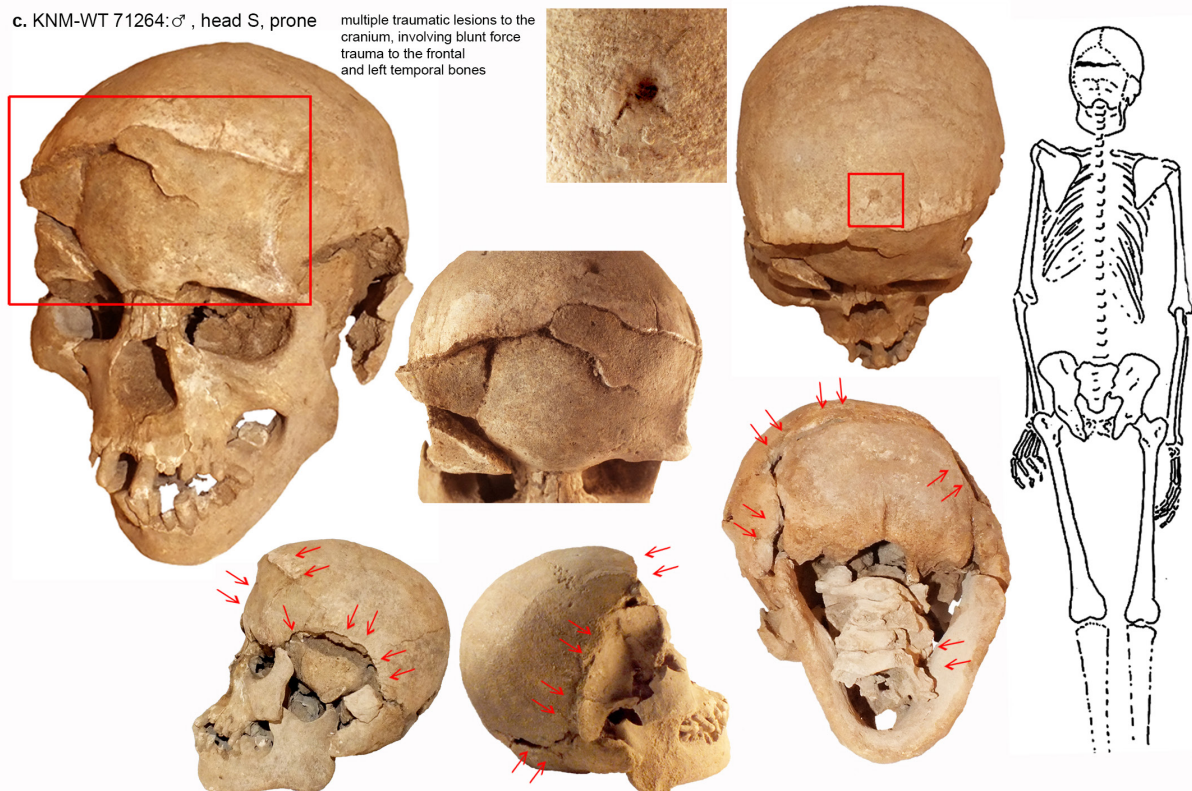
a. KNM-WT 71263: ♂, head NE, face W/down



b. KNM-WT 71265: ♂, head N, face W



c. KNM-WT 71264: ♂, head S, prone



Extended Data Figure 7 | Traumatic lesions on skeletons. a, Skeleton KNM-WT 71263; c, skeleton KNM-WT 71264; b, skeleton KNM-WT 71265. Displayed lesions on KNM-WT 71263 include a linear perforating lesion, with associated depressed fracture on the frontal bone, and a perforation on the right temporal bone. Displayed lesions on KNM-WT 71264 include depressed fractures on the frontal and left parietal bones, and multiple radiating fractures across the skull, as well as a healed (?)

perforating lesion on the frontal bone. Displayed lesions on KNM-WT 71265 include a depressed fracture on the frontal bone. Red boxes indicate enlarged images of particular lesions; red arrows point to the lesions described in Supplementary Information 3.3.10, 3.3.11 and 3.3.12; red dotted line in KNM-WT 71265 delineates what would have been the natural profile of the bone. Photographs and drawings by M.M.L.

Extended Data Table 1 | Dates of sediments, shells and human remains from Nataruk

Relation to human remains Above	Sample	Context	Method	Date 14C	Age	Number	Notes
KNM-WT 72160	Sediment*	Deflated surface	14C	6,440 ± 40 yrs BP	7,430 - 7,270 cal BP	BA-341441	δ13C: -18.6
KNM-WT 72151	Sediment	Deflated surface above skeleton	14C	7,220 ± 40 yrs BP	8,160 - 8,090 cal BP / 8,060 - 7,960 cal BP	BA-344845	δ13C: -19.9
KNM-WT 71251	Sediment Shell**	From excavation	OSL		9,680 ± 805 yrs	X6481	OSL age
		From excavation, adjacent to OSL sample	14C	9,080 ± 35 yrs BP	10,275 - 10,190 cal BP (11,750-9,540 cal BP) ^α	S-ANU#37218	δ13C: -7±2 [‡] 5.1±0.53% calcite
	Solid bone Rib frag	Buried skeleton	U-series		12,200 ± 700 yrs	3503 ANU	
		Buried skeleton	U-series		---	3486 ANU	FAILED [§]
KNM-WT 71252	Solid bone	Buried skeleton	U-series		20,800 ± 1,200 yrs / 21,800 ± 1,400 yrs	3517 ANU	
KNM-WT 71253	Rib frag	Buried skeleton	U-series		12,200 ± 730 yrs / 9,540 ± 860 yrs	3505 ANU	
KNM-WT 71254	Shell Solid bone	From excavation	14C	8,210 ± 40 yrs BP	9,290 - 9,030 cal BP	BA344846	δ13C: +3.5
		Buried skeleton	U-series		6,640 ± 460 yrs	3506 ANU	
KNM-WT 71255	Rib frag	Buried skeleton	U-series		[19,450 ± 4,760 kyr]	3507 ANU	No DAD [¥]
KNM-WT 71257	Rib frag	Buried skeleton	U-series		5,960 ± 550 yrs / 6,670 ± 510 yrs	3508 ANU	
KNM-WT 71258	Shell	Prised from human rib	14C	8,330 ± 40 yrs BP	9,460 - 9,260 cal BP	BA-344847	δ13C: -5.3
	Rib frag	Buried skeleton	U-series		[25,760 ± 3,960 yrs]	3509 ANU	No DAD [¥]
	Rib frag	Buried skeleton	14C		---	BA-343477	FAILED [¥]
KNM-WT 71259	Solid bone	Buried skeleton	U-series		5,590 ± 450 yrs / 6,590 ± 740 yrs	3510 ANU	
	Rib frag	Buried skeleton	14C		---	BA-343478	FAILED [¥]
KNM-WT 71260	Rib frag	Buried skeleton	U-series		[7,190 ± 1,730 yrs]	3511 ANU	No DAD [¥]
	Tooth	Buried skeleton	14C		---	BA-348828	FAILED [¥]
KNM-WT 71263	Rib frag	Buried skeleton	U-series		[18,930 ± 1,540 yrs]	3512 ANU	No DAD [¥]
KNM-WT 71264	Shell	Prised from human skull	14C	9,160 ± 40 yrs BP	10,480 - 10,470 cal BP / 10,420 - 10,230 cal BP	BA-344844	δ13C: -2.5
	Rib frag	Buried skeleton	U-series		[38,990 ± 8,410 yrs]	3513 ANU	No DAD [¥]
	Rib frag	Buried skeleton	U-series		48,300 ± 2,900 yrs	3487 ANU	
KNM-WT 71265	Solid bone	Buried skeleton	U-series		[30,790 ± 6,580 yrs]	3514 ANU	No DAD [¥]
KNM-WT 71266	Solid bone	Buried skeleton	U-series		[7,060 ± 950 yrs]	3515 ANU	No DAD [¥]
KNM-WT 71268	Solid bone	Buried skeleton	U-series		6,680 ± 450 yrs	3504 ANU	
KNM-WT 71274 area	Shell Sediment	Deflated surface	14C	8,450 ± 40 yrs BP	9,530 - 9,430 cal BP	BA-344849	δ13C: -6.9
		Deflated surface	14C		---	BA-344850	FAILED
KNM-WT 71276	Solid bone	Buried skeleton	U-series		8,910 ± 930 yrs	3516 ANU	
Hippopotamus	Solid bone	Surface	U-series		---	3485 ANU	FAILED [§]

* the sample dated is the organic fraction that remains after sieving the sediment to < 180 µm to remove any roots or macrofossils and then acid washed to remove carbonates;

** all dated shells are gastropods;

^α expanded age range taking into account the possible age of the carbon in the ~5% diagenetic calcite observed in the sample;

[‡] δ 13C value is AMS machine-quoted values and was used to correct the age (these can differ from IRMS results);

[§] no relationship between Uranium concentration and apparent age, thus possible Uranium leaching;

[¥] no collagen preserved;

[¥] no DAD results possible, and model conditions violated.

Radiocarbon dates have been calibrated against IntCal13 (ref. 66) in OxCal version 4.2.4 (ref. 65). Radiocarbon age estimates presented as radiocarbon years before the present (years BP) and as calibrated ages (cal. BP) with 95% confidence intervals. OSL and uranium-series age estimates presented in years (yrs). Error on the OSL age is at 1σ; errors on the uranium-series ages are at 2σ.

Extended Data Table 2 | Chronological synthesis and uranium-series sample characterization

Relation to human remains	U-series sample characterisation	Discussion	Chronological Synthesis
Surface sediments (above skeletons KNM-WT 71260 & 71251)		14C dates from surface sediments (7,430–7,270 cal BP; 8,160–8,090/8,060–7,960 cal BP); expected and observed younger than the human remains	7,270 – 8,160 yrs
KNM-WT 71251	3486: A piece of thick but porous bone; sample contains mineral crusts. Data obtained along a single track. The U concentrations are low (0.1–2.35 ppm), with U/Th ratios 0.2–67, indicating a high degree of incorporation of detrital Th, compromising any U-series age calculations. 3503: A solid piece of cortical bone. Data obtained along two tracks. There is Th contamination close to the outside of the 1 st track, while the 2 nd is characterised by much lower U concentrations and low U/Th ratios. The remainder of the sample yields a DAD age of 12,200 ± 700 yrs.	Consistent 14C (10,275–10,190 cal BP), OSL (9,680 ± 805 yrs) and U-series dates (12,200 ± 700 yrs), in direct sedimentary or depositional relation to the skeleton and on the remains themselves	9,680 – 12,200 yrs
KNM-WT 71252	3517: A bone with a thick cortical section. Data obtained along two tracks. Both sets are contaminated at the outside with Th. DAD results: 20,800 ± 1,200 and 21,800 ± 1,400 yrs.	U-series date, with possible U leaching	
KNM-WT 71253	3505: A rib with relatively thick cortical sections. Data collected along four lines (1–7, 8–13, 14–19, 20–24). DAD results only on the 1 st two sections: 12,200 ± 700 and 9,500 ± 900 yrs.	U-series date only (minimum age)	> 9,500 – 12,200 yrs
KNM-WT 71254	3506: Long bone fragment with increasing pores towards the inner part. Data collected along a single line. There is some Th contamination at the outside and at spot analysis 16. DAD result: 6,640 ± 460 yrs.	14C date (9,290–9,040 cal BP) on shell associated with human skeleton	9,030 – 9,290 yrs [> 6,600 yrs]
KNM-WT 71255	3507: A rib with very thin cortical bone. Data collected along four lines (1–7, 8–10, 11–15, 16–19). Many analyses show low U-concentrations combined with high Th contamination; no DAD result. Average individual age estimates is 19,450 yrs, with a standard deviation of 4,760 yrs.	U-series date with no DAD results, and thus violated model conditions	
KNM-WT 71257	3508: A rib with thick cortical bone. Data collected along four lines (1–6, 7–19, 20–25, 26–29); DAD results on 1 st and 2 nd sets: 5,960 ± 550 and 6,670 ± 510 yrs.	U-series date only (minimum age)	> 6,000 yrs
KNM-WT 71258	3509: A rib with thicker cortical bone. Data collected along two lines (1–10, 11–15); no DAD result on the 1 st data set. The average of the individual age estimates is 25,760 yrs, with a standard deviation of 3,960 yrs.	14C date on shell prised from human skeleton (9,260–9,460 cal BP); U-series age with no DAD results and violated model conditions	9,260 – 9,460 yrs
KNM-WT 71259	3510: Bone with relatively thin cortical section. Data collected along 2 lines (1–12, 13–24). In the 1 st set, sampling spots towards the centre of the bone were highly contaminated with Th. DAD results: 6,590 ± 740 and 5,590 ± 450 yrs.	U-series date, on thin cortical bone (minimum age)	> 5,600 yrs
KNM-WT 71260	3511: A rib with thick cortical section. Data collected along 4 lines (1–4, 5–9, 10–14, 15–19). No Th contamination. No DAD result possible. Average individual age estimates: 7,190 yrs, with a standard deviation of 1,730 yrs.	U-series date that yielded no DAD results and thus violated model conditions	
KNM-WT 71263	3512: A rib with thick cortical bone. Data collected along 2 lines. There is significant Th contamination, and only 2 data points pass the reliability criteria, indicating an age of ~18,900 yrs.	U-series date that yielded no DAD results and thus violated model conditions	
KNM-WT 71264	3487: Bone fragment with a thick cortical section. Data obtained along 1 track. Low U concentrations (< 1.1 ppm), the outer analyses interfered by Th. The inner data yield a DAD age of 48,300 ± 2,900 yrs. 3513: Rib with thick cortical bone on 1 side (first 2 lines: 1–5, 6–15) and nearly no cortical bone on the other (last 2 lines 16–20, 21–26); some Th contamination at the outside as well as from pores (#13, #14). No DAD result possible. Average individual age estimates: 38,990 yrs, with a standard deviation of 8,410 yrs.	14C date on shell prised from human skeleton (10,480–10,470/10,420–10,230 cal BP); 2 U-series age estimations that yielded consistently old ages – one yielded no DAD results and thus violated model conditions; the other may show U leaching	10,230 – 10,480 yrs [? 39,000 – 48,300 yrs]
KNM-WT 71265	3514: Solid piece of cortical bone. Data collected along 1 line. All analysis yielded relatively low U concentrations, but virtually no Th. No DAD result possible. Average individual age estimates: 30,790 yrs, with a standard deviation of 6,580 yrs.	U-series date that yielded no DAD results and thus violated model conditions	
KNM-WT 71266	3515: Solid piece of cortical bone. Data collected along 1 line. Virtually no Th. No DAD result possible. Average individual age estimates: 7,060 yrs, with a standard deviation of 950 yrs.	U-series date that yielded no DAD results and thus violated model conditions	
KNM-WT 71268	3504: Long bone fragment with increasing pores towards the inner part. Data obtained along 1 track. One analysis is contaminated by sediments in a pore. DAD age: 6,680 ± 450 yrs.	U-series date only (minimum age)	> 6,700 yrs
KNM-WT 71274 area		14C on shell from among a scatter of deflated human bones (9,540–9,430 cal BP)	9,430 – 9,530 yrs
KNM-WT 71276	3516: Solid bone with spongy section. Data collected along 1 line. Only 1 analysis is contaminated with Th (#8). DAD result: 8,920 ± 900 yrs.	U-series date only (minimum age)	> 8,900 yrs
Hippopotamus	3485: Solid fragment of hippopotamus cortical bone. Data obtained along 1 track. Dark black bone that has a very different preservation to the human bones; sample is highly mineralised and shows extremely low U/Th elemental ratios (0.1–12) throughout, rendering any U-series age calculation irrelevant.		

Radiocarbon age estimates presented as calibrated ages (cal BP) with 95% confidence intervals. OSL and U-series age estimates presented in years (yrs). Error on the OSL age is at 1 σ ; errors on the U-series ages are at 2 σ .

Motor neurons control locomotor circuit function retrogradely via gap junctions

Jianren Song¹, Konstantinos Ampatzis¹, E. Rebecka Björnfors¹ & Abdeljabbar El Manira¹

Motor neurons are the final stage of neural processing for the execution of motor behaviours. Traditionally, motor neurons have been viewed as the ‘final common pathway’, serving as passive recipients merely conveying to the muscles the final motor program generated by upstream interneuron circuits^{1,2}. Here we reveal an unforeseen role of motor neurons in controlling the locomotor circuit function via gap junctions in zebrafish. These gap junctions mediate a retrograde analogue propagation of voltage fluctuations from motor neurons to control the synaptic release and recruitment of the upstream V2a interneurons that drive locomotion. Selective inhibition of motor neurons during ongoing locomotion de-recruits V2a interneurons and strongly influences locomotor circuit function. Rather than acting as separate units, gap junctions unite motor neurons and V2a interneurons into functional ensembles endowed with a retrograde analogue computation essential for locomotor rhythm generation. These results show that motor neurons are not a passive recipient of motor commands but an integral component of the neural circuits responsible for motor behaviour.

The existence of electrical coupling between motor neurons and the premotor excitatory V2a interneurons was examined using three complementary sets of experiments in the adult zebrafish. First, we show the occurrence of dye coupling via gap junctions from motor neurons to V2a interneurons following the injection of neurobiotin into muscles (Fig. 1a) or intracellularly into single motor neurons (Fig. 1b). However, practically no dye coupling was seen between motor neurons and the excitatory commissural (V0v) interneurons or the inhibitory glycinergic interneurons (Fig. 1c–f). Second, we reveal, using immunohistochemistry, that connexin 35/36 was co-localized with close appositions between motor neuron dendrites and V2a interneuron axons (Fig. 1g–i). Moreover, connexin 35/36 punctae on motor neuron dendrites were co-localized with most of the synaptophysin-labelled presynaptic terminals (Fig. 1h). Finally, the occurrence of bidirectional electrical coupling was confirmed electrophysiologically using whole-cell paired recordings from V2a interneurons and motor neurons ($n = 57$ pairs, 57 zebrafish, Fig. 1j–l). We previously showed that V2a interneurons are subdivided into three sub-classes that are selectively connected to slow, intermediate and fast motor neurons and form three separate modules³. Electrical coupling occurred between pairs of the same module as well as between pairs of adjacent modules (Fig. 1m).

Electrical coupling occurred only between V2a interneuron and motor neuron pairs connected by chemical synapses. In these pairs, stimulation of the presynaptic V2a interneurons elicited monosynaptic excitatory postsynaptic potentials (EPSPs) in the postsynaptic motor neurons that were blocked with ionotropic glutamate antagonists or cadmium, and unmasked EPSP-like events mediated by electrical coupling (Extended Data Fig. 1a, d). Conversely, stimulation of the motor neurons also elicited fast electrical events in the presynaptic V2a interneurons that were insensitive to ionotropic glutamate receptor antagonists or cadmium, indicating that they are mediated by

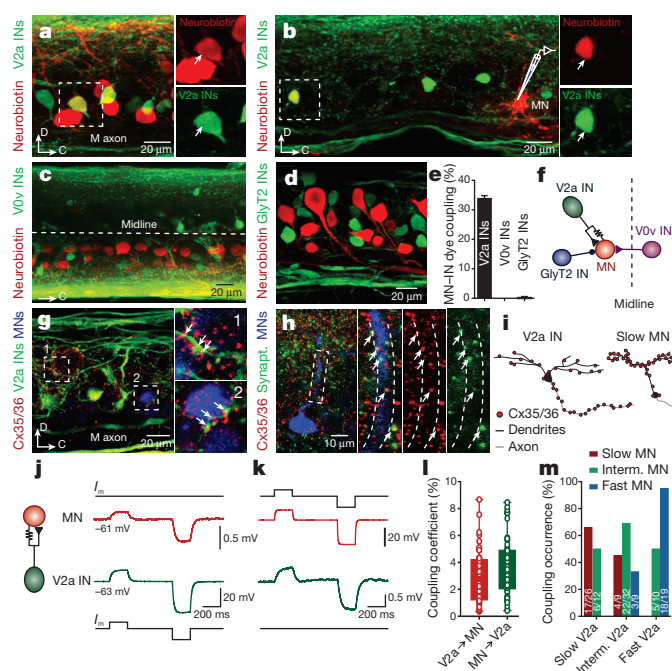


Figure 1 | Existence of gap junctions between V2a interneurons and motor neurons. **a**, Neurobiotin injections into the axial muscles of the transgenic *Chx10:GFP* line labelled motor neurons (MNs) and electrically coupled V2a interneurons (INs). The Mauthner (M) axon is indicated. **b**, Neurobiotin injection into a single motor neuron (red) resulted in dye coupling of V2a interneurons. **c**, No dye coupling was observed between motor neurons and excitatory commissural V0v interneurons. **d**, No dye coupling was seen between motor neurons and inhibitory (GlyT2) interneurons. **e**, Percentage of V2a, GlyT2, and V0v interneurons dye-coupled with motor neurons. **f**, Diagram showing interneurons connected to motor neurons with mixed or chemical synapses. **g**, Cx35/36 punctae were found on V2a interneuron axons that made close contacts with motor neuron dendrites. Higher magnifications of the dashed boxes in **g** are shown on the right. **h**, Cx35/36 (red) punctae on dendrites of neurobiotin labelled motor neurons (blue) making close contacts with presynaptic terminals expressing synaptophysin (green). **i**, Cx35/36 distribution in V2a interneurons and slow/intermediate motor neurons. **j**, **k**, Representative recordings showing bidirectional electrical coupling between V2a interneurons and motor neurons ($n = 57$ pairs from 57 zebrafish). **l**, Coupling coefficient (percentage of voltage changes transferred between electrically coupled neurons) in both directions (V2a interneurons to motor neurons = $2.8 \pm 0.2\%$ and motor neuron to V2a interneurons = $3.5 \pm 0.3\%$, $n = 57$). **m**, Percentage of occurrence of electrical coupling between synaptically connected V2a interneuron subclasses and slow, intermediate and fast motor neurons. The occurrence of electrical coupling was significantly higher between pairs of the same module ($76.7 \pm 9.2\%$, $n = 77$) than between pairs of adjacent modules ($29.7 \pm 9.7\%$, $n = 40$) ($P < 0.01$, two-tailed Student's *t*-test).

¹Department of Neuroscience, Karolinska Institute, 171 77 Stockholm, Sweden.

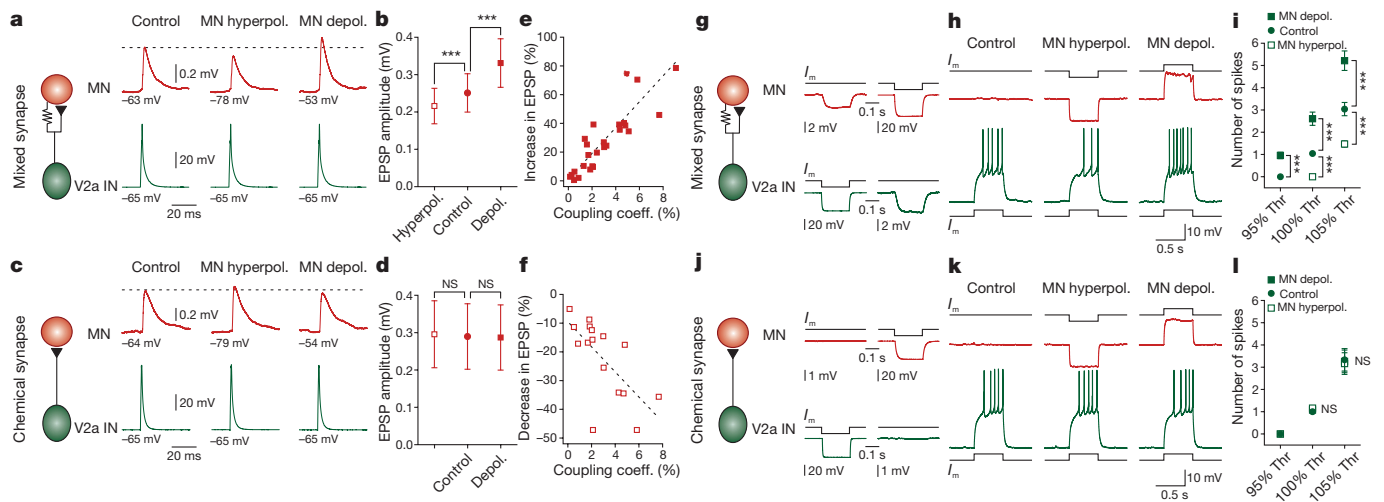


Figure 2 | Motor neurons control the strength of synaptic transmission and the firing threshold of V2a interneurons. **a**, Representative recording of a V2a interneuron and a motor neuron connected with mixed synapses ($n = 26$ pairs from 26 zebrafish). Strength of synaptic transmission of the V2a interneuron was strongly influenced by the motor neuron membrane potential. **b**, EPSP amplitude as a function of the motor neuron membrane potential (control, 0.25 ± 0.05 mV; hyperpolarization, 0.21 ± 0.05 mV; depolarization, 0.33 ± 0.07 mV; $n = 26$, $P < 0.001$, two-tailed Student's t -test, error bars denote s.e.m.). **c**, Representative recording of a V2a interneuron and a motor neuron only connected with chemical synapses ($n = 8$ pairs from 8 zebrafish). Strength of the V2a interneuron synaptic transmission was not affected by the motor neuron membrane potential. **d**, Motor neuron membrane potential changes had no effect on the EPSP amplitude (control, 0.29 ± 0.09 mV; hyperpolarization, 0.29 ± 0.09 mV; depolarization, 0.28 ± 0.08 mV; $n = 8$, $P > 0.05$, two-tailed Student's t -test, error bars denote s.e.m.). **e**, Correlation ($R^2 = 0.74$) between the EPSP amplitude increase and the coupling coefficient. **f**, Correlation

electrical coupling (Extended Data Fig. 1a–f). In pairs connected only with chemical synapses, monosynaptic EPSPs elicited in motor neurons by V2a interneurons were completely blocked by ionotropic glutamate receptor antagonists or cadmium. In these pairs, however, there was no effect in the backward direction from the motor neurons to the V2a interneurons (Extended Data Fig. 1g–i).

Anatomical analysis showed that connexin 35/36 was co-localized at sites of close appositions between V2a interneuron axons and motor neuron dendrites of the pairs connected with mixed synapses (Extended Data Fig. 2a–c), but it was absent in all close contact sites of the pairs connected only with chemical synapses (Extended Data Fig. 2d–f). Furthermore, there was no spatial segregation between the pairs connected with mixed electrical and chemical synapses and those only connected with chemical synapses (Extended Data Fig. 2g). While in the latter pairs the monosynaptic EPSP was abolished by ionotropic glutamate receptor antagonists or cadmium, an electrical component always remained in pairs connected with mixed synapses (Extended Data Fig. 2h). Together the above results show that gap junctions occur between presynaptic terminals of V2a interneurons and dendrites of motor neurons exclusively in pairs connected with chemical synapses. This enables backward propagation of electrical signals from motor neurons to the upstream excitatory V2a interneurons.

In addition to their mixed electrical and chemical connections with V2a interneurons, motor neurons are also electrically coupled to each other (Extended Data Fig. 3a–c). Electrically coupled motor neurons display dendro-dendritic close appositions co-localized with connexin 35/36 punctae (Extended Data Fig. 3d, e). Furthermore, dye coupling also occurred between V2a interneurons. Neurobiotin injection into single V2a interneurons resulted in dye coupling of other V2a interneurons (Extended Data Fig. 3f). Thus, electrical coupling exists

($R^2 = 0.45$) between the EPSP amplitude decrease and the coupling coefficient. **g**, Representative recording of a V2a interneuron and a motor neuron connected with mixed synapses ($n = 23$ pairs from 23 zebrafish). **h**, V2a interneuron firing was strongly influenced by the motor neuron membrane potential. **i**, Influence of motor neuron membrane potential on V2a interneuron firing ($n = 23$, $P < 0.001$, two-tailed Student's t -test; filled circle, control; open square, motor neuron hyperpolarization; filled square, motor neuron depolarization; error bars denote s.e.m.). **j**, Representative recording of a V2a interneuron and a motor neuron connected only via chemical synapses ($n = 12$ from 12 zebrafish). **k**, V2a interneuron firing was not affected by changing the motor neuron membrane potential. **l**, Motor neuron membrane potential had no effect on V2a interneuron firing in pairs connected by chemical synapses ($n = 12$, $P > 0.05$, two-tailed Student's t -test; filled circle, control; open square, motor neuron hyperpolarization; filled square, motor neuron depolarization; error bars denote s.e.m.).

among motor neurons and V2a interneurons in addition to between these two neuronal populations.

Next, we examined if the backward propagation of electrical signals from motor neurons can influence, in an analogue fashion, the V2a interneurons. For this, we first assessed if changes in motor neuron membrane potentials affected the synaptic strength (Fig. 2a–f) and firing threshold (Fig. 2g–l) of V2a interneurons. Notably, a steady change in the motor neuron membrane potentials substantially changed the amplitude of the monosynaptic EPSPs in the pairs connected with gap junctions (Fig. 2a, b). There was a strong correlation between the coefficient of electrical coupling and the changes in the EPSP amplitude induced by the depolarization or hyperpolarization of motor neurons (Fig. 2e, f). Importantly, in the pairs connected only by chemical synapses, the monosynaptic EPSP amplitude was not affected by steady changes in motor neuron membrane potentials (Fig. 2c, d).

We then tested if the backward propagation of electrical signals via gap junctions can extend the influence of motor neurons to the V2a interneuron somata to change their firing threshold. Indeed, in pairs connected with gap junctions (Fig. 2g), a steady hyperpolarization of the motor neurons significantly increased the V2a interneuron firing threshold and decreased their repetitive firing (Fig. 2h, i). Conversely, a steady depolarization of motor neurons decreased the V2a interneuron firing threshold and increased their repetitive firing (Fig. 2h, i). In contrast, in pairs connected only with chemical synapses (Fig. 2j), changing the motor neuron's membrane potential had no effect on the V2a interneuron's threshold and firing activity (Fig. 2k, l). Thus, gap junctions enable backward propagation of analogue electrical signals from motor neurons to the V2a interneurons that contributes to defining the strength of their synaptic transmission and firing threshold.

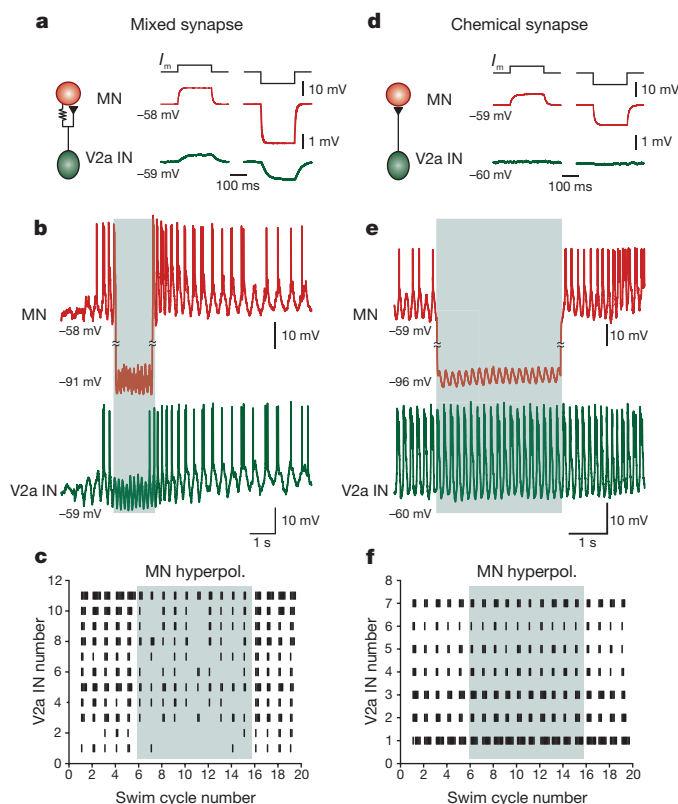


Figure 3 | Motor neurons control the recruitment of V2a interneurons during locomotion. **a**, Representative recording of a V2a interneuron and a motor neuron connected with mixed synapses ($n = 11$ pairs from 11 zebrafish). **b**, In the same pair as **a**, both the V2a interneuron and the motor neuron were rhythmically active and firing action potentials during swimming-related activity. Steady hyperpolarization of the motor neuron (from -62.5 ± 0.6 mV to -97.5 ± 1.6 mV, $n = 11$) reversibly inhibited the V2a interneuron firing. **c**, The number of action potentials per swimming-related cycle in the V2a interneurons was reduced by hyperpolarization of the motor neurons in pairs connected with mixed synapses (average number of action potentials before: 3.1 ± 0.1 , during: 1.0 ± 0.1 , and after the motor neuron hyperpolarization: 3.0 ± 0.2) (before versus during: $P < 0.0001$, $n = 11$, two-tailed Student's *t*-test). **d**, Representative recording of a V2a interneuron and a motor neuron connected only with chemical synapses ($n = 7$ from 7 zebrafish). **e**, Hyperpolarization of the motor neurons (from -61.5 ± 0.7 mV to -101.5 ± 2.3 mV, $n = 7$) had no effect on the swimming-related firing of the V2a interneuron. **f**, The number of action potentials per swimming-related cycle in the V2a interneurons was not affected in the by hyperpolarization of the motor neurons in pairs only connected with chemical synapses (average number of action potentials before: 3.4 ± 0.2 , during: 3.3 ± 0.1 , and after the motor neuron hyperpolarization: 3.3 ± 0.1) (before versus during: $P > 0.05$, $n = 7$, two-tailed Student's *t*-test).

V2a interneurons play an important role in generating locomotor rhythm and driving motor neuron activity⁴. While motor neurons have traditionally been excluded from contributing to rhythm generation, their strong influence on V2a interneurons prompted us to reassess their role in the locomotor central pattern generator. Whole-cell recordings were made from V2a interneuron and motor neuron pairs connected with mixed (Fig. 3a–c) or only with chemical synapses (Fig. 3d–f). Both the V2a interneurons and the motor neurons were rhythmically active during a locomotor episode induced by stimulation of descending projections from the brainstem. Remarkably, in pairs connected with gap junctions, a steady hyperpolarization of the motor neurons reversibly suppressed the V2a interneuron activity (Fig. 3b). Hyperpolarization of the motor neurons strongly and reversibly decreased the number of action potentials in V2a interneurons during each locomotor cycle (Fig. 3c). By contrast, in pairs only connected

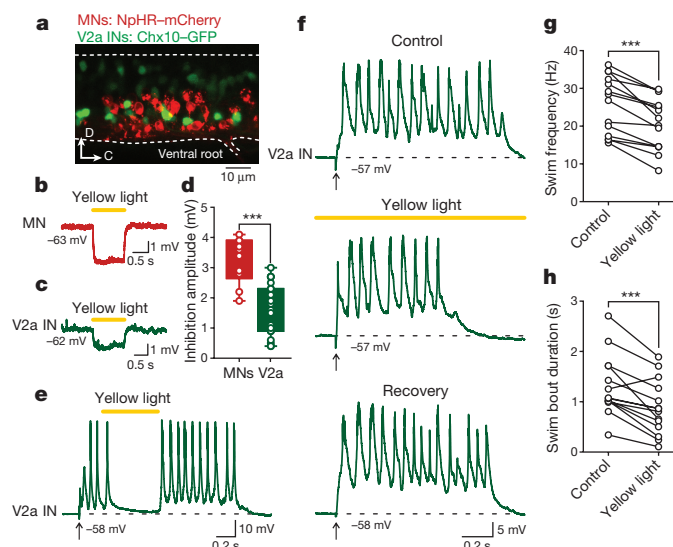


Figure 4 | Optogenetic inhibition of motor neurons decreases the frequency of locomotion. **a**, A transgenic zebrafish line which expressed halorhodopsin (NpHR-mCherry) in motor neurons and GFP in V2a interneurons was used. **b**, **c**, Representative recording of a motor neuron ($n = 10$) and V2a interneuron ($n = 15$) that were hyperpolarized by yellow light stimulation. **d**, Box and whisker plot showing that yellow-light-induced inhibition was larger ($P < 0.001$, two-tailed Student's *t*-test, bars denote minimum and maximum values) in motor neurons than V2a interneurons. **e**, Yellow light stimulation over two spinal segments reversibly stopped the rhythmic swimming-related activity in a representative V2a interneuron ($n = 5$ of 19 from 19 zebrafish). **f**, In other preparations ($n = 14$ of 19), similar yellow light stimulation decreased the burst frequency and duration of the swimming-related activity in a representative V2a interneuron. **g**, Optogenetic inhibition of motor neurons decreased the swimming-related burst frequency (control, 25.7 ± 2.0 Hz; yellow light, 20.6 ± 1.8 Hz; $n = 14$, $P < 0.001$, two-tailed Student's *t*-test). **h**, Optogenetic inhibition of motor neurons decreased the swimming-related activity duration (control, 1.31 ± 0.16 s; yellow light, 0.88 ± 0.14 s; $n = 14$, $P < 0.001$, two-tailed Student's *t*-test).

with chemical synapses (Fig. 3d), hyperpolarization of the motor neurons did not affect the activity of the V2a interneurons (Fig. 3e, f). These results uncover a prominent role of motor neurons, extending their influence retrogradely to control the recruitment threshold and firing activity of the V2a interneurons—a pivotal interneuron component of the rhythm generator circuit.

Finally, we examined if acute inhibition of motor neurons during ongoing locomotor activity alters features of the locomotor rhythm generation. To this end, we used transgenic zebrafish in which halorhodopsin (NpHR-mCherry) expression was targeted to motor neurons, but not to V2a interneurons, which only expressed GFP (Fig. 4a). Optogenetic stimulation of halorhodopsin with yellow light over two spinal segments hyperpolarized the motor neurons (Fig. 4b). This hyperpolarization also propagated to the V2a interneurons via gap junctions (Fig. 4c, d). Moreover, during ongoing locomotor activity, optogenetic inhibition of motor neurons reversibly de-recruited some V2a interneurons ($n = 5$ of 19 preparations, Fig. 4e). In the other 14 preparations, optogenetic inhibition of motor neurons substantially decreased the burst frequency (by 20%, Fig. 4f, g) and the duration (by 33%, Fig. 4f, h) of the locomotor activity. None of these effects were induced by the yellow light in control zebrafish without NpHR-mCherry expression (Extended Data Fig. 4).

Motor neurons are traditionally considered passive recipients of motor programs generated by upstream interneuron circuits. Here we reveal an unforeseen role of motor neurons that extends their influence to premotor excitatory interneurons and embeds them within the core circuit generating the locomotor rhythm. Thus far, the only established route for motor neuron influence in the spinal cord was via Renshaw

cells, which produce recurrent inhibition but do not contribute to the rhythm generation^{5–8}. We now demonstrate a novel and direct path via gap junctions that mediates a simultaneous integration of ongoing activity of motor neurons by upstream excitatory interneurons. The existence of electrical coupling between motor neurons and the excitatory V2a interneurons suggests a major revision of the construction of locomotor circuits. The phasic change of the membrane potential of motor neurons during each locomotor cycle is immediately transferred to upstream excitatory interneurons and affects their recruitment and synaptic release. Rather than acting only as feed-forward units, the bidirectional transfer of electrical signals via electrical coupling ties motor neurons and interneurons into functional ensembles endowed with an analogue processing essential for the elaboration of the locomotor rhythm. Indeed, motor neurons serve more complex functions and might use additional mechanisms in conjunction with gap junctions to influence the function of neural circuits for motor behaviour.

Mixed synapses are widespread in vertebrates^{9,10}, including mammals. Indeed, there is growing evidence of the presence of glutamatergic mixed electrical and chemical synapses in the mammalian nervous system^{11–14}. An important function ascribed to electrical coupling via gap junctions is to synchronize the activity of neuronal ensembles^{15,16}. However, retrograde intercellular communication via gap junctions has been shown in sensory systems of fish and crustaceans, and can lead to synchronization of the activity of sensory axons^{17–23}. In addition, gap junctions can synchronize the activity of interneurons and motor neurons to produce rhythmic activity^{24–30}. Anatomical substrates exist in the adult mammalian spinal cord^{11,12} to enable retrograde influence from motor neurons onto the premotor locomotor circuit as we have now revealed in the adult zebrafish. Therefore we propose that motor neurons have a strong and immediate influence on the locomotor circuit and are not merely passive recipients of motor commands but rather an integral component of the spinal circuit generating the locomotor rhythm.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 June; accepted 27 November 2015.

Published online 13 January 2016.

- Liddell, E. G. T. & Sherrington, C. S. Vol. 97 488–518 (Proceedings of the Royal Society of London, 1925).
- Sherrington, C. S. Vol. 97 519–545 (Proceedings of the Royal Society of London, 1925).
- Ampatzis, K., Song, J., Ausborn, J. & El Manira, A. Separate microcircuit modules of distinct V2a interneurons and motoneurons control the speed of locomotion. *Neuron* **83**, 934–943 (2014).
- El Manira, A. Dynamics and plasticity of spinal locomotor circuits. *Curr. Opin. Neurobiol.* **29**, 133–141 (2014).
- Mentis, G. Z. *et al.* Noncholinergic excitatory actions of motoneurons in the neonatal mammalian spinal cord. *Proc. Natl Acad. Sci. USA* **102**, 7344–7349 (2005).
- Nishimaru, H., Restrepo, C. E., Ryge, J., Yanagawa, Y. & Kiehn, O. Mammalian motor neurons corelease glutamate and acetylcholine at central synapses. *Proc. Natl Acad. Sci. USA* **102**, 5245–5249 (2005).
- Noga, B. R., Shefchyk, S. J., Jamal, J. & Jordan, L. M. The role of Renshaw cells in locomotion: antagonism of their excitation from motor axon collaterals with intravenous mecamylamine. *Exp. Brain Res.* **66**, 99–105 (1987).
- Pratt, C. A. & Jordan, L. M. Ia inhibitory interneurons and Renshaw cells as contributors to the spinal mechanisms of fictive locomotion. *J. Neurophysiol.* **57**, 56–71 (1987).
- Pereda, A. E. Electrical synapses and their functional interactions with chemical synapses. *Nature Rev. Neurosci.* **15**, 250–263 (2014).
- Serrano-Velez, J. L. *et al.* Abundance of gap junctions at glutamatergic mixed synapses in adult Mosquitofish spinal cord neurons. *Front. Neural Circuits* **8**, 66 (2014).

- Bautista, W., McCrea, D. A. & Nagy, J. I. Connexin36 identified at morphologically mixed chemical/electrical synapses on trigeminal motoneurons and at primary afferent terminals on spinal cord neurons in adult mouse and rat. *Neuroscience* **263**, 159–180 (2014).
- Rash, J. E. *et al.* Mixed synapses discovered and mapped throughout mammalian spinal cord. *Proc. Natl Acad. Sci. USA* **93**, 4235–4239 (1996).
- Hamzei-Sichani, F. *et al.* Mixed electrical-chemical synapses in adult rat hippocampus are primarily glutamatergic and coupled by connexin-36. *Front. Neuroanat.* **6**, 13 (2012).
- Vivar, C., Traub, R. D. & Gutierrez, R. Mixed electrical-chemical transmission between hippocampal mossy fibers and pyramidal cells. *Eur. J. Neurosci.* **35**, 76–82 (2012).
- Bennett, M. V. & Zukin, R. S. Electrical coupling and neuronal synchronization in the Mammalian brain. *Neuron* **41**, 495–511 (2004).
- Connors, B. W. & Long, M. A. Electrical synapses in the mammalian brain. *Annu. Rev. Neurosci.* **27**, 393–418 (2004).
- Curti, S. & Pereda, A. E. Voltage-dependent enhancement of electrical coupling by a subthreshold sodium current. *J. Neurosci.* **24**, 3999–4010 (2004).
- El Manira, A., Cattaert, D., Wallen, P., DiCaprio, R. A. & Clarac, F. Electrical coupling of mechanoreceptor afferents in the crayfish: a possible mechanism for enhancement of sensory signal transmission. *J. Neurophysiol.* **69**, 2248–2251 (1993).
- Herberholz, J., Antonsen, B. L. & Edwards, D. H. A lateral excitatory network in the escape circuit of crayfish. *J. Neurosci.* **22**, 9078–9085 (2002).
- Korn, H., Sotelo, C. & Bennett, M. V. L. The lateral vestibular nucleus of the toadfish *Opus tau*: ultrastructural and electrophysiological observations with special reference to electronic transmission. *Neuroscience* **2**, 851–884 (1977).
- Pereda, A. E., Bell, T. D. & Faber, D. S. Retrograde synaptic communication via gap junctions coupling auditory afferents to the Mauthner cell. *J. Neurosci.* **15**, 5943–5955 (1995).
- Slesinger, P. & Bell, C. C. Primary afferent fibers conduct impulses in both directions under physiological stimulus conditions. *J. Comp. Physiol.* **157**, 15–22 (1985).
- Zipser, B. & Bennett, M. V. Interaction of electrosensory and electromotor signals in lateral line lobe of a mormyrid fish. *J. Neurophysiol.* **39**, 713–721 (1976).
- Personius, K. E., Chang, Q., Mentis, G. Z., O'Donovan, M. J. & Balice-Gordon, R. J. Reduced gap junctional coupling leads to uncorrelated motor neuron firing and precocious neuromuscular synapse elimination. *Proc. Natl Acad. Sci. USA* **104**, 11808–11813 (2007).
- Li, W. C., Roberts, A. & Sofke, S. R. Locomotor rhythm maintenance: electrical coupling among premotor excitatory interneurons in the brainstem and spinal cord of young *Xenopus* tadpoles. *J. Physiol. (Lond.)* **587**, 1677–1693 (2009).
- Rekling, J. C., Shao, X. M. & Feldman, J. L. Electrical coupling and excitatory synaptic transmission between rhythmogenic respiratory neurons in the preBötzinger complex. *J. Neurosci.* **20**, RC113 (2000).
- Saint-Amant, L. & Drapeau, P. Motoneuron activity patterns related to the earliest behavior of the zebrafish embryo. *J. Neurosci.* **20**, 3964–3972 (2000).
- Sillar, K. T. & Simmers, A. J. Electrical coupling and intrinsic neuronal oscillations in *Rana temporaria* spinal cord. *Eur. J. Morphol.* **32**, 293–298 (1994).
- Tresch, M. C. & Kiehn, O. Motor coordination without action potentials in the mammalian spinal cord. *Nature Neurosci.* **3**, 593–599 (2000).
- Zhang, H. Y., Li, W. C., Heitler, W. J. & Sillar, K. T. Electrical coupling synchronises spinal motoneuron activity during swimming in hatchling *Xenopus* tadpoles. *J. Physiol. (Lond.)* **587**, 4455–4466 (2009).

Acknowledgements We thank C. Wyart, S. Grillner, O. Kiehn and G. Silberberg as well as the members of our lab for comments on an early version of the manuscript. We are grateful to A. Baradel, C. Wyart, S. Higashijima and H. Baier for providing the zebrafish lines used in this study. This study was supported by grants from the Swedish Research Council, Karolinska Institute and the Swedish Brain Foundation.

Author Contributions J.S., K.A. and A.E.M. initiated the project and designed the experiments. J.S. performed all the electrophysiological and optogenetic experiments. K.A. performed the anatomical and contributed to the electrophysiological experiments. E.R.B. contributed to the anatomical and electrophysiological experiments. All the authors contributed to data analysis, discussed the results and participated in writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.E.M. (Abdel.ElManira@ki.se).

METHODS

The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Zebrafish (*Danio rerio*) were raised and kept in a core facility at the Karolinska Institute according to established procedures. All experimental protocols were approved by the local Animal Research Ethical Committee, Stockholm and were performed in accordance with EU guidelines.

Adult zebrafish (8–10 weeks old) of either sex were anaesthetized in 0.03% tricaine methane sulfonate (MS-222, Sigma-Aldrich). Retrograde labelling of axial motor neurons was performed using dye injections with Rhodamine-dextran (3000 MW, Molecular Probes) into muscles or with dextran Alexa Fluor 647 (10000 MW, Molecular probes)³¹. To reveal possible dye coupling between V2a interneurons and motor neurons, the low molecular weight (287 Da) intracellular tracer neurobiotin (25%, Vector Laboratories) was injected into the muscles of the transgenic zebrafish lines: Tg[*Chx10:GFP*] in which expression of GFP in the V2a interneurons was driven by the promoter *Chx10* (ref. 32), Tg[*glt2:GFP*] in which GFP is expressed in inhibitory interneurons or by crossing Tg[*vglut2a:loxP-DsRed-loxP-GFP*] with Tg[*dbx1b:Cre*] to drive the expression of GFP in the commissural V0v interneurons^{33,34}. All the injected animals were allowed to recover overnight to help the retrograde transport of the tracer.

All animals were deeply anaesthetized with 0.1% MS-222 and intracardially perfused with 4% paraformaldehyde (PFA) in phosphate buffer (PB, 0.1 M, pH = 7.4). The spinal cord was dissected out and post-fixed in 4% PFA solution overnight at 4°C. For the dye coupling experiments tissue was processed for single immunohistochemical labelling against GFP. The spinal cords were washed three times for 5 min in phosphate buffered saline (0.01 M PBS, pH = 7.4). Non-specific protein binding sites were blocked with 0.15% normal horse serum with 5% bovine serum albumin (BSA) and 1% Triton X-100 in PBS for 30 min at room temperature. Spinal cords were then incubated with rabbit anti-GFP (A11122, Life Techn., dilution 1:500) in 1% Triton X-100 in PBS at 4°C for 72 h. After thorough buffer rinses the tissue was incubated with anti-rabbit Alexa Fluor 488 antibody at 1:500 dilution overnight at 4°C and subsequently rinsed in PBS and coverslipped with anti-fade fluorescent mounting medium (Vectashield Hard Set, Vector Labs). To reveal the exact location of gap junctions between motor neurons and V2a interneurons and the possible co-localization with chemical synapses, mouse anti-connexin 35/36 (MAB3045, Millipore) was applied at 1:200 dilution along with rabbit anti-GFP (A11122, dilution 1:500) or with rabbit anti-synaptophysin (ab23754, Abcam, dilution 1:100). Whole-mount imaging of the spinal cords was carried out in a laser scanning confocal microscope (LSM 510 Meta, Zeiss).

Eight- to ten-week-old adult zebrafish were used in the electrophysiological experiments. Animals were anaesthetized using MS-222, which was dissolved in frozen extracellular solution before the dissection of the brainstem and spinal cord together with the spinal column and a piece of tail muscle^{35–37}. The dissected preparation was then transferred to a recording chamber maintained at room temperature of 20–22°C that was continuously perfused with an extracellular solution containing (in mM): 134 NaCl, 2.9 KCl, 2.1 CaCl₂, 1.2 MgCl₂, 10 HEPES and 10 glucose with pH 7.8 adjusted with NaOH and an osmolarity of 290 mOsm. Muscle contractions were blocked by adding D-tubocurarine (10 µM, Sigma-Aldrich) to the extracellular solution.

An extracellular stimulation electrode was placed at the border between the brainstem and spinal cord to induce swimming-related activity (10 pulses at 1 or 2 Hz). Identified motor neurons pre-labelled with a retrograde tracer and GFP-positive V2a interneurons were recorded using whole-cell patch-clamp technique. The patch-clamp electrodes were filled with a solution contained (in mM): 120 K-gluconate, 5 KCl, 10 HEPES, 4 Mg₂ATP, 0.3 Na₄GTP, 10 Na-phosphocreatine with pH 7.4 adjusted with KOH and an osmolarity of 275 mOsm. Neurons were visualized using a fluorescence microscope (Axioskop FS Plus, Zeiss) equipped with IR-differential interference contrast (DIC) optics and a CCD camera. Electrical signals were amplified (MultiClamp 700B; Molecular Devices) and low-pass filtered at 10 kHz. Paired recordings were made within the same segment from randomly targeted motor neurons and V2a interneurons belonging to the slow or intermediate microcircuit module. Single and multiple current pulses (0.5–1 ms at 40 Hz) were used to elicit action potentials in either V2a interneurons or motor neurons to record the resulting synaptic events in motor neurons or V2a interneurons. The occurrence of electrical coupling was tested using long-duration depolarizing and hyperpolarizing current pulses in V2a interneurons and motor neurons. The glutamatergic nature of the excitatory postsynaptic potentials (EPSPs) was determined by using ionotropic glutamate receptors antagonists (AP-5 and NBQX).

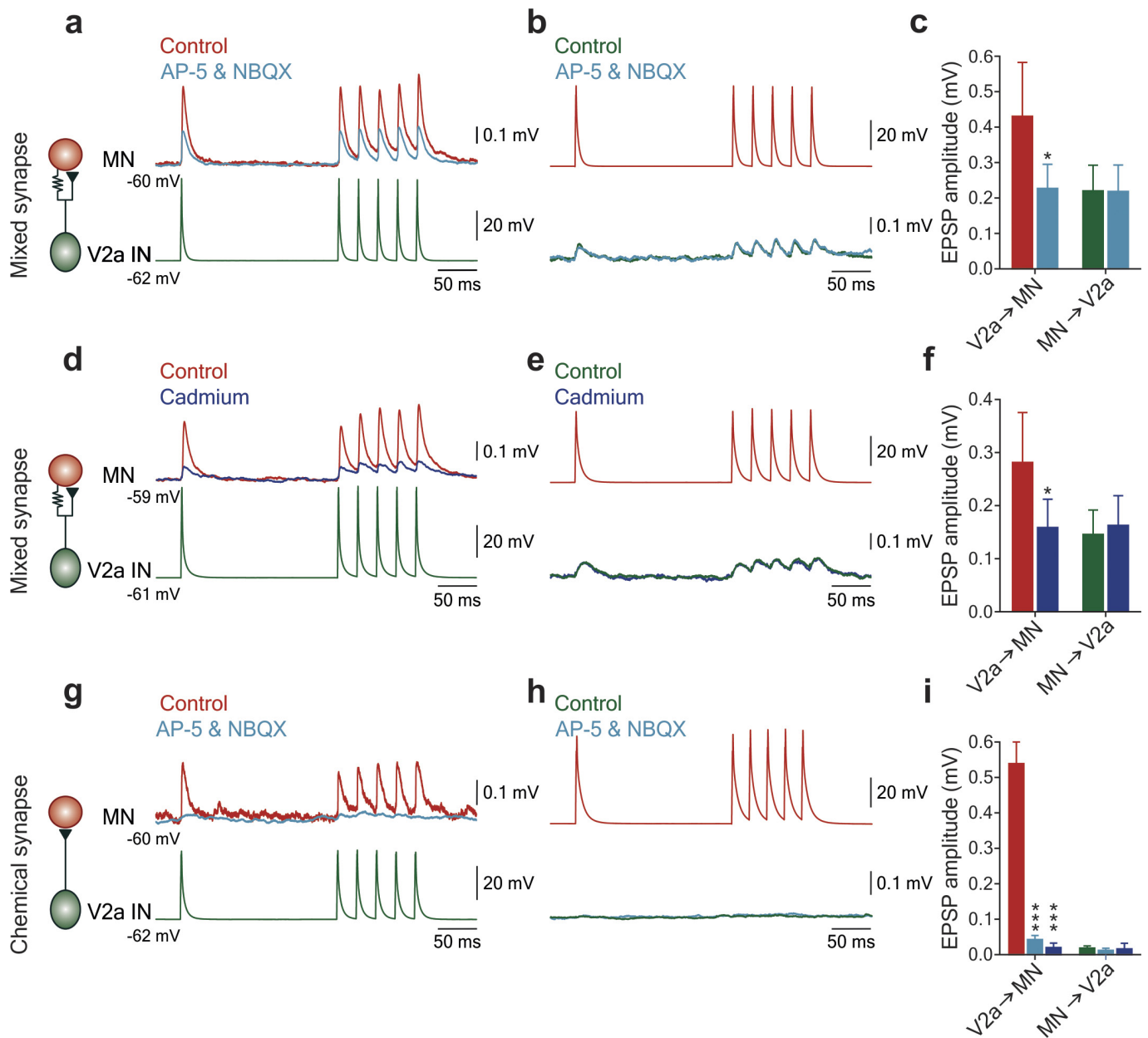
In some experiments single motor neurons or V2a interneurons were passively filled with 2% neurobiotin to examine the occurrence of dye coupling. Furthermore, pairs of recorded V2a interneurons and motor neurons were also passively filled

with 0.25% neurobiotin for post hoc analysis of their morphology. Spinal cords with neurobiotin-filled neurons were dissected out and transferred into 4% PFA solution overnight at 4°C. The tissue was then washed extensively with PBS and incubated in streptavidin conjugated to Alexa Fluor 488 (1:500, Invitrogen) or Alexa Fluor 555 (1:500, Invitrogen) and were used for detection of connexin 35/36 immunoreactivity following the protocol described above. The occurrence of close appositions between neuronal elements and their co-localization with connexin 35/36 were determined from single optical sections acquired with laser scanning confocal microscope (LSM 510 Meta, Zeiss). These high magnification sections were filtered and pseudo-coloured for optimal visualization of structures from different neurons. The neuronal processes (axons and dendrites) of each labelled neuron were then traced and reconstructed using semi-automated tracing software (NeuronJ, plugin of ImageJ).

For optogenetic experiment, we used a zebrafish line Gal4^{s1020t} (Et(-0.6hsp70l:Gal4-VP16)s1020t), in which Gal4-VP16 was expressed in motor neurons^{38,39}. Crossing the Gal4^{s1020t} line with the Tg[UAS:NpHR-mCherry]^{s1989t} (refs 40, 41) line yielded zebrafish expressing of NpHR-mCherry in motor neurons. By crossing these zebrafish with another zebrafish line Tg[*Chx10:GFP*], expressing of GFP in the V2a interneurons, we finally obtained zebrafish simultaneously expressing NpHR-mCherry in motor neurons and GFP in V2a interneurons. Gal4^{s1020t} also expresses Gal4 in Kolmer-Agduhr cells, which are GABAergic cerebrospinal-fluid-contacting cells that are not active in immobilized animals³⁹. To avoid any effects from these cells, we blocked GABA_A receptors with gabazine in these experiments. The NpHR-mCherry expression declined in adult zebrafish, we therefore used 3–4-week-old animals in optogenetic experiments. The specific expression of NpHR-mCherry was confirmed by injecting a retrograde fluorescent dye in the muscle. In addition, we ensured that there was no expression of NpHR-mCherry in the GFP-labelled V2a interneurons. Motor neurons were inhibited optogenetically in two spinal segments by using light stimulation (yellow light: 580 nm), which was generated by a custom-made LED system⁴². The zebrafish line Tg[*Chx10:GFP*] was used in control experiments to test the effects of yellow light stimulation in animals without NpHR-mCherry. Because the ontogenetic stimulation protocol was the same in all conditions, the investigators were not blinded to whether zebrafish expressed, or did not, NpHR-mCherry.

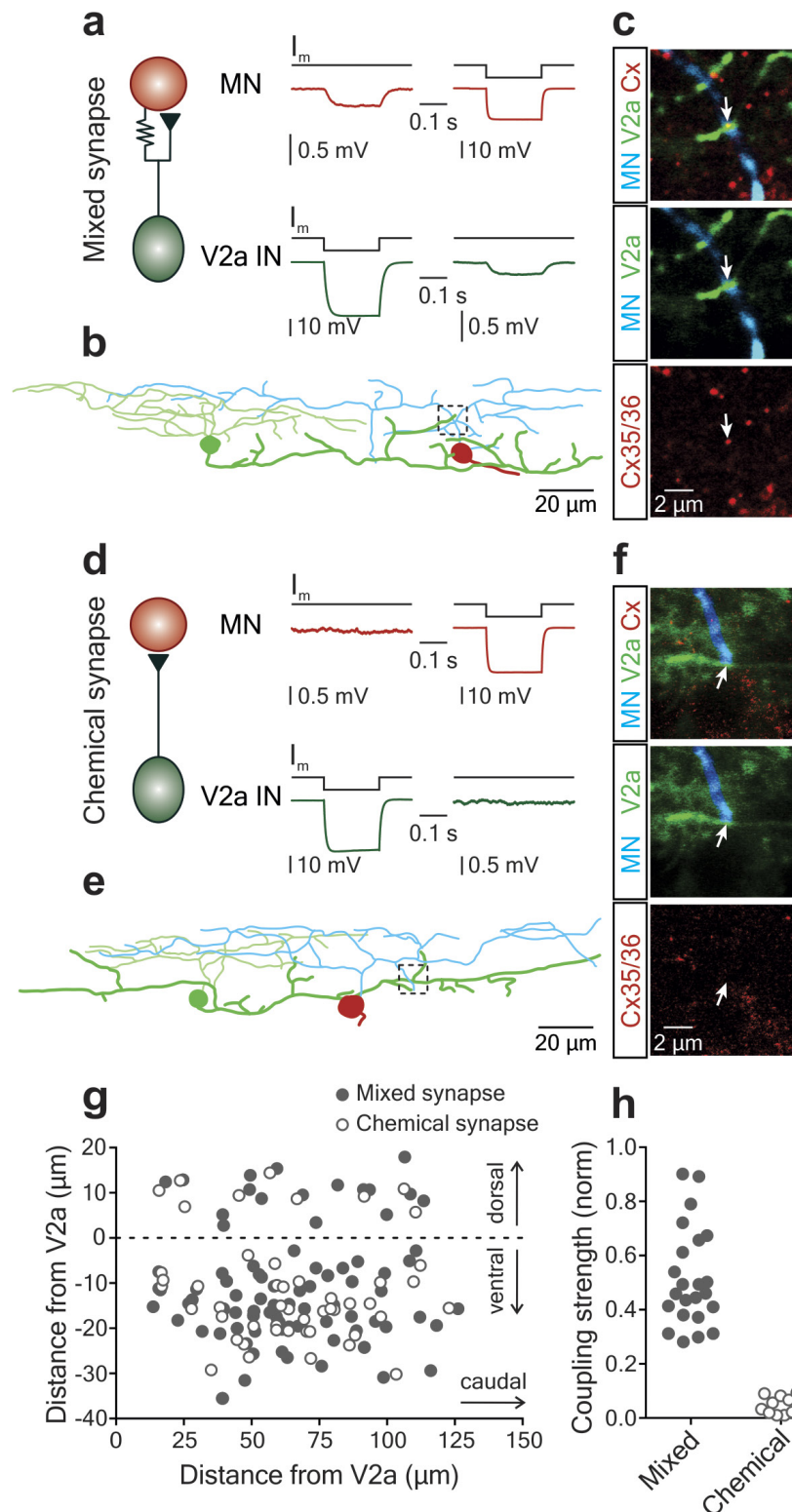
The analysis of the electrophysiological data was performed using Spike2 (Cambridge Electronic Design) or Clampfit (Molecular Devices) software. The dual recording traces illustrated represent averages of 60–200 consecutive sweeps. The EPSP amplitude was calculated as the difference between baseline and peak. No statistical methods were used to predetermine sample size. The sample size was determined based on numbers ($n \geq 6$ from different animals) reported in previous studies and no samples were excluded. The significance of differences between the means in experimental groups and conditions was analysed using two-tailed Student's *t*-test (Prism, GraphPad Software Inc.). The homogeneity of variances was tested using an *F*-test and was found to be similar between sample groups. Differences were considered to be significant if $P < 0.05$. All data presented here are given as mean \pm s.e.m. and only one data point (n) per animal was used.

1. Ampatzis, K., Song, J., Ausborn, J. & El Manira, A. Pattern of innervation and recruitment of different classes of motoneurons in adult zebrafish. *J. Neurosci.* **33**, 10875–10886 (2013).
2. Kimura, Y., Okamura, Y. & Higashijima, S. *alx*, a zebrafish homolog of *Chx10*, marks ipsilateral descending excitatory interneurons that participate in the regulation of spinal locomotor circuits. *J. Neurosci.* **26**, 5684–5697 (2006).
3. Satou, C., Kimura, Y. & Higashijima, S. Generation of multiple classes of V0 neurons in zebrafish spinal cord: progenitor heterogeneity and temporal control of neuronal diversity. *J. Neurosci.* **32**, 1771–1783 (2012).
4. Satou, C. *et al.* Transgenic tools to characterize neuronal properties of discrete populations of zebrafish neurons. *Development* **140**, 3927–3931 (2013).
5. Ausborn, J., Mahmood, R. & El Manira, A. Decoding the rules of recruitment of excitatory interneurons in the adult zebrafish locomotor network. *Proc. Natl Acad. Sci. USA* **109**, E3631–E3639 (2012).
6. Gabriel, J. P. *et al.* Principles governing recruitment of motoneurons during swimming in zebrafish. *Nature Neurosci.* **14**, 93–99 (2011).
7. Kyriakatos, A. *et al.* Initiation of locomotion in adult zebrafish. *J. Neurosci.* **31**, 8422–8431 (2011).
8. Scott, E. K. *et al.* Targeting neural circuitry in zebrafish using GAL4 enhancer trapping. *Nature Methods* **4**, 323–326 (2007).
9. Wyart, C. *et al.* Optogenetic dissection of a behavioural module in the vertebrate spinal cord. *Nature* **461**, 407–410 (2009).
10. Arrenberg, A. B., Del Bene, F. & Baier, H. Optical control of zebrafish behavior with halorhodopsin. *Proc. Natl Acad. Sci. USA* **106**, 17968–17973 (2009).
11. Schoonheim, P. J., Arrenberg, A. B., Del Bene, F. & Baier, H. Optogenetic localization and genetic perturbation of saccade-generating neurons in zebrafish. *J. Neurosci.* **30**, 7111–7120 (2010).
12. Ljunggren, E. E., Haupt, S., Ausborn, J., Ampatzis, K. & El Manira, A. Optogenetic activation of excitatory premotor interneurons is sufficient to generate coordinated locomotor activity in larval zebrafish. *J. Neurosci.* **34**, 134–139 (2014).



Extended Data Figure 1 | Co-existence of electrical and chemical synapses. **a**, Representative recording of a V2a interneuron and a motor neuron connected with mixed synapses ($n = 11$ from 11 zebrafish). EPSP amplitude induced in the motor neuron by stimulation of the V2a interneuron was reduced by blockers of ionotropic glutamate receptors (AP-5 and NBQX) leaving only an electrical event. **b**, Stimulation of the motor neuron also induced an electrical event in the V2a interneuron that was insensitive to the glutamate receptor blockers. **c**, Changes in the amplitude of the synaptic events in motor neurons and V2a interneurons induced by glutamate receptor blockers ($*P < 0.05$, two-tailed Student's t -test, error bars denote s.e.m.). **d**, Representative recording of a V2a interneuron and a motor neuron ($n = 12$ from 12 zebrafish) in which blockade of chemical synaptic transmission with cadmium decreased the

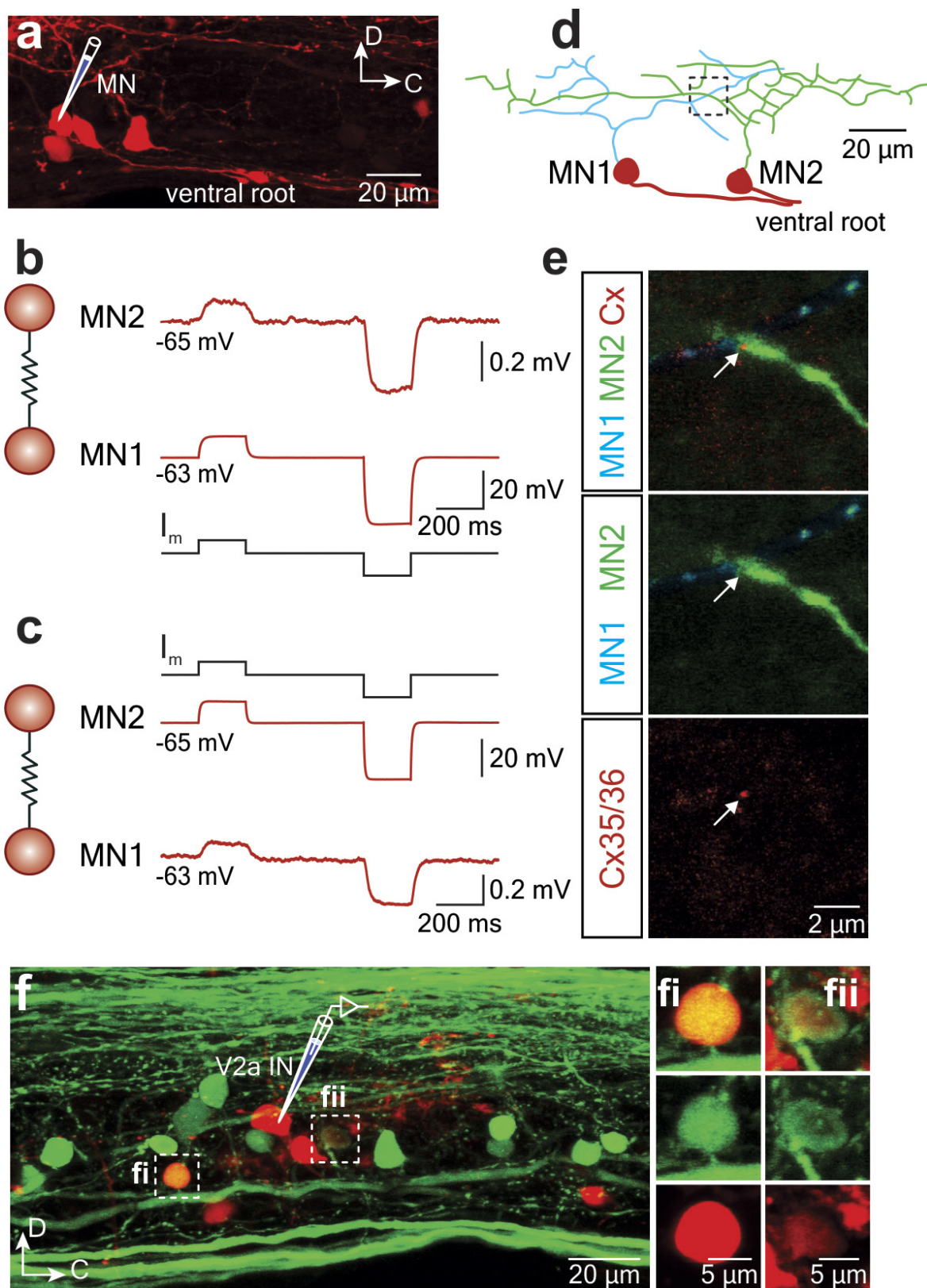
EPSP amplitude leaving only an electrical event. **e**, Cadmium did not affect the amplitude of the electrical events induced in the V2a interneuron by stimulation of the motor neuron. **f**, Effect of cadmium on the synaptic events in motor neurons and V2a interneurons ($*P < 0.05$, two-tailed Student's t -test, error bars denote s.e.m.). **g**, Representative recording of a V2a interneuron and a motor neuron connected with chemical synapses ($n = 11$ from 11 zebrafish). Blockade of ionotropic glutamate receptors completely abolished the V2a interneuron-induced EPSP in the motor neuron. **h**, In this pair, stimulation of the motor neuron did not induce any electrical event in the V2a interneuron. **i**, Effects of glutamate blockers and cadmium on the synaptic events induced in motor neurons and V2a interneurons ($***P < 0.001$, two-tailed Student's t -test, error bars denote s.e.m.).



Extended Data Figure 2 | Co-localization of connexin 35/36 with putative synaptic sites between V2a interneurons and motor neurons.

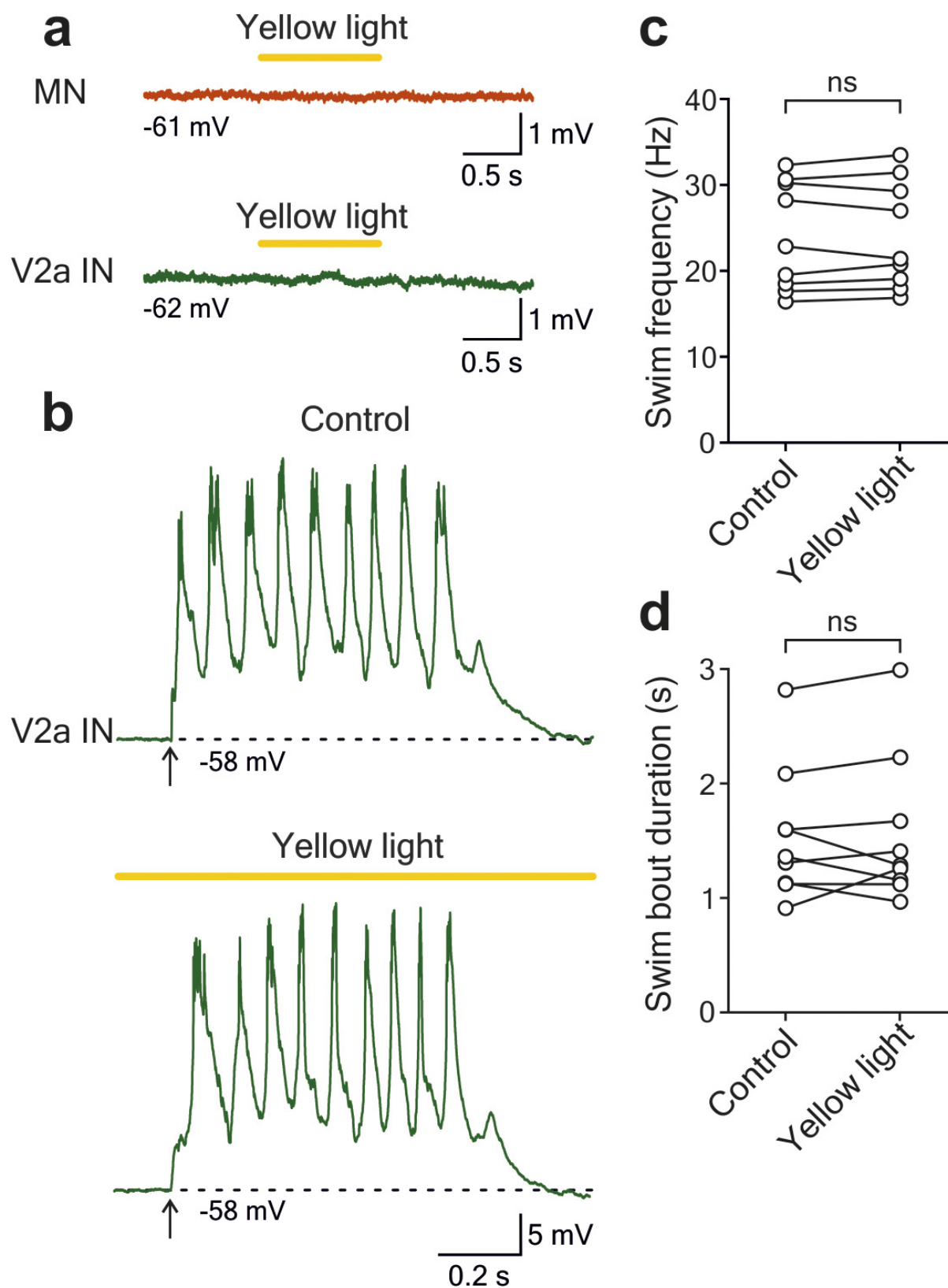
a, Representative recording of a V2a interneuron and a motor neuron connected with mixed synapses that were filled with neurobiotin ($n = 7$ from 7 zebrafish). **b**, Reconstruction of the recorded neurons revealed a zone of close apposition between V2a interneuron axon collaterals and motor neuron dendrites. **c**, Connexin 35/36 (Cx) was co-localized with close appositions between V2a interneuron axons and motor neuron dendrites. **d**, Representative recording of a V2a interneuron and a motor neuron connected only with chemical synapses that were filled with neurobiotin ($n = 3$ from 3 zebrafish). **e**, Reconstruction of the recorded

pair revealed close appositions between V2a interneuron axon collaterals and motor neuron dendrites. **f**, In pairs only connected with chemical synapses no co-localization of connexin 35/36 was observed with close appositions between V2a interneuron axons and motor neuron dendrites. **g**, The position of motor neuron cell bodies relative to those of the V2a interneurons (0 position on both axes) in pairs connected with mixed or only with chemical synapses. **h**, Scatter plot showing the coupling strength (electrical component/EPSP amplitude) in pairs connected with mixed or only with chemical synapses. The amplitude of the electrical component was determined after blockade of chemical synaptic transmission with ionotropic glutamate receptor antagonists or cadmium.



Extended Data Figure 3 | Electrical coupling between motor neurons. **a**, Injection of neurobiotin into a single motor neuron (red) resulted in dye coupling of other motor neurons. **b**, **c**, Representative recording of a pair of motor neurons (MN1 and MN2) displaying bidirectional electrical coupling ($n = 11$ from 11 zebrafish). **d**, Reconstruction of the recorded

motor neuron pair in **b** and **c** revealed a zone of dendro-dendritic close apposition. **e**, Connexin 35/36 (Cx) co-localized very closely to the motor neuron dendrites. **f**, Injection of neurobiotin into a single V2a interneuron (red) resulted in dye coupling of other V2a interneurons ($n = 6$ from 6 zebrafish). **fi**, **fii**, Higher magnifications of the dashed boxes in **f**.



Extended Data Figure 4 | Yellow light had no effect in control animals without NpHR expression. **a**, Representative recordings showing that yellow light stimulation did not induce any change in the membrane potential of the motor neurons and the V2a interneurons. **b**, Representative recording of a V2a interneuron from a control zebrafish without NpHR-mCherry expression. Yellow light stimulation over two

spinal segments had no effect on the rhythmic swimming-related activity in the V2a interneuron ($n = 9$ from 9 zebrafish). **c**, Plot showing the lack of effect of the yellow light stimulation on the swimming-related activity frequency ($n = 9$, $P > 0.05$, two-tailed Student's t -test). **d**, Graph showing the lack of effect of the yellow light on the swimming-related activity duration ($n = 9$, $P > 0.05$, two-tailed Student's t -test).

NANOG alone induces germ cells in primed epiblast *in vitro* by activation of enhancers

Kazuhiro Murakami^{1,2,3,4,5*}, Ufuk Günesdogan^{1,2,3*}, Jan J. Zyllicz^{1,2,3}, Walfred W. C. Tang^{1,2,3}, Roopsha Sengupta^{1,2,3}, Toshihiro Kobayashi^{1,2,3}, Shinseog Kim^{1,2,3}, Richard Butler¹, Sabine Dietmann³ & M. Azim Surani^{1,2,3}

Nanog, a core pluripotency factor in the inner cell mass of blastocysts, is also expressed in unipotent primordial germ cells (PGCs) in mice¹, where its precise role is yet unclear^{2–4}. We investigated this in an *in vitro* model, in which naive pluripotent embryonic stem (ES) cells cultured in basic fibroblast growth factor (bFGF) and activin A develop as epiblast-like cells (EpiLCs) and gain competence for a PGC-like fate⁵. Consequently, bone morphogenetic protein 4 (BMP4), or ectopic expression of key germline transcription factors *Prdm1*, *Prdm14* and *Tfap2c*, directly induce PGC-like cells (PGCLCs) in EpiLCs, but not in ES cells^{6–8}. Here we report an unexpected discovery that *Nanog* alone can induce PGCLCs in EpiLCs, independently of BMP4. We propose that after the dissolution of the naive ES-cell pluripotency network during establishment of EpiLCs^{9,10}, the epigenome is reset for cell fate determination. Indeed, we found genome-wide changes in NANOG-binding patterns between ES cells and EpiLCs, indicating epigenetic resetting of regulatory elements. Accordingly, we show that NANOG can bind and activate enhancers of *Prdm1* and *Prdm14* in EpiLCs *in vitro*; BLIMP1 (encoded by *Prdm1*) then directly induces *Tfap2c*. Furthermore, while SOX2 and NANOG promote the pluripotent state in ES cells, they show contrasting roles in EpiLCs, as *Sox2* specifically represses PGCLC induction by *Nanog*. This study demonstrates a broadly applicable mechanistic principle for how cells acquire competence for cell fate determination, resulting in the context-dependent roles of key transcription factors during development.

Transcription factors and epigenetic changes confer competence for somatic and PGC fates when naive pluripotent inner cell mass (ICM) from embryonic day (E)3.5–4.5 blastocysts develop to primed epiblast at ~E6.0 (ref. 11). Similarly, naive pluripotent ES cells in 2i medium acquire competency within ~48 h after culture in bFGF and activin A *in vitro*, when day-2 EpiLCs differentiate into PGCLCs in response to BMP4 (ref. 5). These putative PGCLCs show expression of Δ PE-Oct3/4-GFP (hereafter termed GOF-GFP) and *Prdm1*-GFP reporters (Fig. 1a and Extended Data Fig. 1a–c) after upregulation of the three key regulators of PGCLCs: *Prdm1* (encoding BLIMP1), *Prdm14*, and *Tfap2c* (encoding AP-2 γ)^{5,7,8}.

NANOG and PRDM14 share similar binding profiles in ES cells and contribute to pluripotency¹². While *Prdm14* is also a key regulator of PGC fate^{13,14}, the role of *Nanog* is unclear, although *Nanog* is detected in E6.5 posterior proximal epiblast^{15,16}, the site of PGC induction, and thereafter in the early germ line^{1,7}. However, we unexpectedly found that doxycycline (Dox) induced expression of *Nanog* alone stimulated GOF-GFP and *Prdm1*-GFP expression in day-2 EpiLCs, indicating specification of putative PGCLCs (Fig. 1a and Extended Data Figs 1a, d–f, 2a–e). Furthermore, *Nanog* apparently acts synergistically with BMP4 to increase the number of GFP⁺ cells, which we

did not see with Oct3/4 (Extended Data Fig. 2f–h). *Nanog* induced PGCLCs in the presence of noggin, a BMP signalling inhibitor, demonstrating that it acts independently of BMP–SMAD signalling (Fig. 1b). Physiological (equivalent to ES cells) or higher levels of NANOG induced PGCLCs with similar efficiency (Extended Data Fig. 3a–c).

We analysed *Nanog*-induced GFP⁺ cells, sorted by fluorescence-activated cell sorting (FACS), which showed upregulation of the key PGC regulators, *Prdm1*, *Prdm14* and *Tfap2c*, as well as *Nanos3* and *Dppa3*, but downregulation of ES-cell-specific *Klf4* (Fig. 1c and Extended Data Fig. 3d–f). This mirrors the response seen with BMP4-mediated PGCLC induction⁵. Notably, principal component analysis (PCA) of global gene expression confirmed that *Nanog*- and BMP4-induced day-4 PGCLCs are highly similar, and closely match with the previously reported day-6 PGCLCs⁵ (Fig. 1d and Extended Data Fig. 3g–j). Furthermore, BLIMP1, PRDM14 and AP-2 γ (but not KLF4) were detected in PGCLCs by immunofluorescence (Fig. 1e and Extended Data Fig. 4). Thus, *Nanog* clearly induces PGC-like fate in EpiLCs and not their reversion to ES cells.

The *Nanog*-induced PGCLCs also showed unique early germline-specific epigenetic modifications: global enrichment of H3K27me3 and erasure of H3K9me2 (Fig. 1e and Extended Data Fig. 4)^{17,18}, together with the initiation of DNA demethylation through the repression of *Uhrf1*, *Dnmt3a* and *Dnmt3b* (Fig. 1c and Extended Data Fig. 3e, i), and upregulation of 5-hydroxymethylcytosine (5hmC) and TET1 (Extended Data Fig. 4)¹⁹. Expression of *Dazl* also indicated progression of DNA demethylation in PGCLCs (Extended Data Fig. 4a, b), which is reminiscent of BMP4-induced PGCLCs⁵.

Next, we asked if *Nanog*- and cytokine-induced PGCLCs could dedifferentiate into pluripotent embryonic germ cells (EGCs), as seen with E8.5 PGCs¹¹. We first subjected PGCLCs to a selection with retinoic acid and bFGF for 5 days, which promotes PGCLC proliferation, but not of ES-cell-like cells^{20,21} (Extended Data Fig. 5a, b). The resulting PGCLCs were transferred to 2i/LIF medium to promote their dedifferentiation into EGC-like cells (EGCLCs), which after several passages produced self-renewing GFP⁺ EGCLCs (Extended Data Fig. 5b). These EGCLCs when introduced into blastocysts contributed extensively (in 27/29 embryos) to chimaeric fetuses at E9.5 (Extended Data Fig. 5c), unlike ‘unipotent’ PGCLCs/PGCs, which neither integrated nor contributed to the fetus (Extended Data Fig. 5d–g).

We then sought genetic evidence that *Nanog* induces bona fide PGCLCs using ES cells with a mutation in *Prdm1*, which is obligatory for PGC specification, but not for the pluripotent state^{22,23}. Consistently, no PGCLCs were induced from *Prdm1*^{−/−} day-2 EpiLCs, nor did they revert to ES cells (Fig. 2a and Extended Data Fig. 6a, b). Instead, the aggregates showed somatic gene expression, including *Hoxa1* and *Hoxb1*, which is reminiscent of the aborted PGC fate in

¹Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK. ²Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK. ³Wellcome Trust Medical Research Council Stem Cell Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK. ⁴Laboratory for Pluripotent Cell Studies, Center for Developmental Biology, RIKEN, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan. ⁵Laboratory for Molecular and Cellular Biology, Faculty of Advanced Life Science, Hokkaido University, Kita21 Nishi11, Kita-ku, Sapporo, Hokkaido 001-0021, Japan.

*These authors contributed equally to this work.

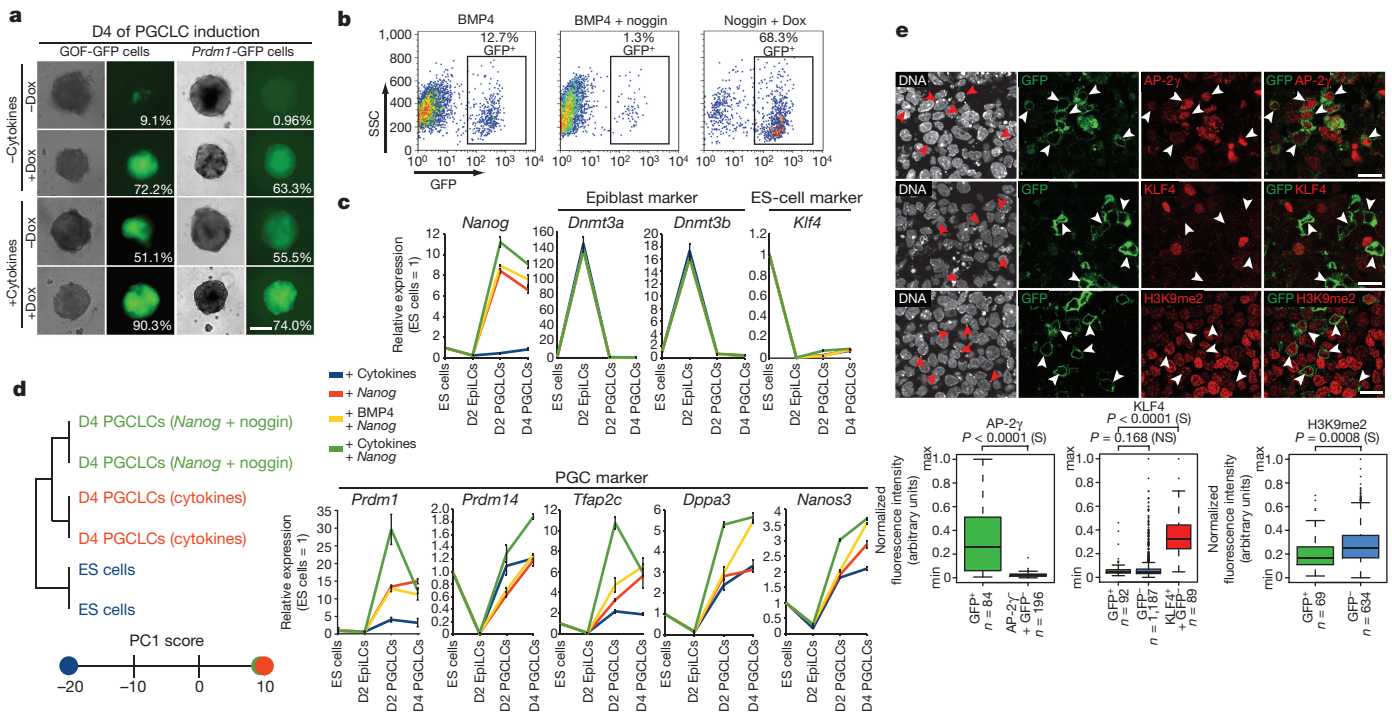


Figure 1 | *Nanog* induces PGCLCs in EpiLCs. **a**, Brightfield/GFP images representing day-4 male PGCLCs induced by *Nanog* (+Dox); percentages are GFP⁺ cells after FACS. Scale bar, 200 μ m. **b**, FACS plots for GOF-GFP⁺ day-4 PGCLCs induced by BMP4 or *Nanog* (+Dox) and with or without noggin. SSC, side scatter. **c**, Analysis of male GOF-GFP cells (quantitative polymerase chain reaction (qPCR)) as indicated. GFP⁺ cells were FACS sorted. $\Delta\Delta C_t \pm$ standard deviation (s.d.); $n = 3$ biological

replicates. **d**, Microarray analyses of GOF-GFP ES cells and day-4 PGCLCs; unsupervised hierarchical clustering and principal component (PC)1 scores. **e**, Immunofluorescence of *Nanog*-induced BLIMP1-GFP⁺ PGCLCs at day 4. Arrowheads highlight single GFP⁺ cells; two biological replicates. Scale bar, 10 μ m. Two-sided/unpaired t -test; NS, not significant ($P > 0.01$); S, significant ($P \leq 0.01$); $n =$ number of cells.

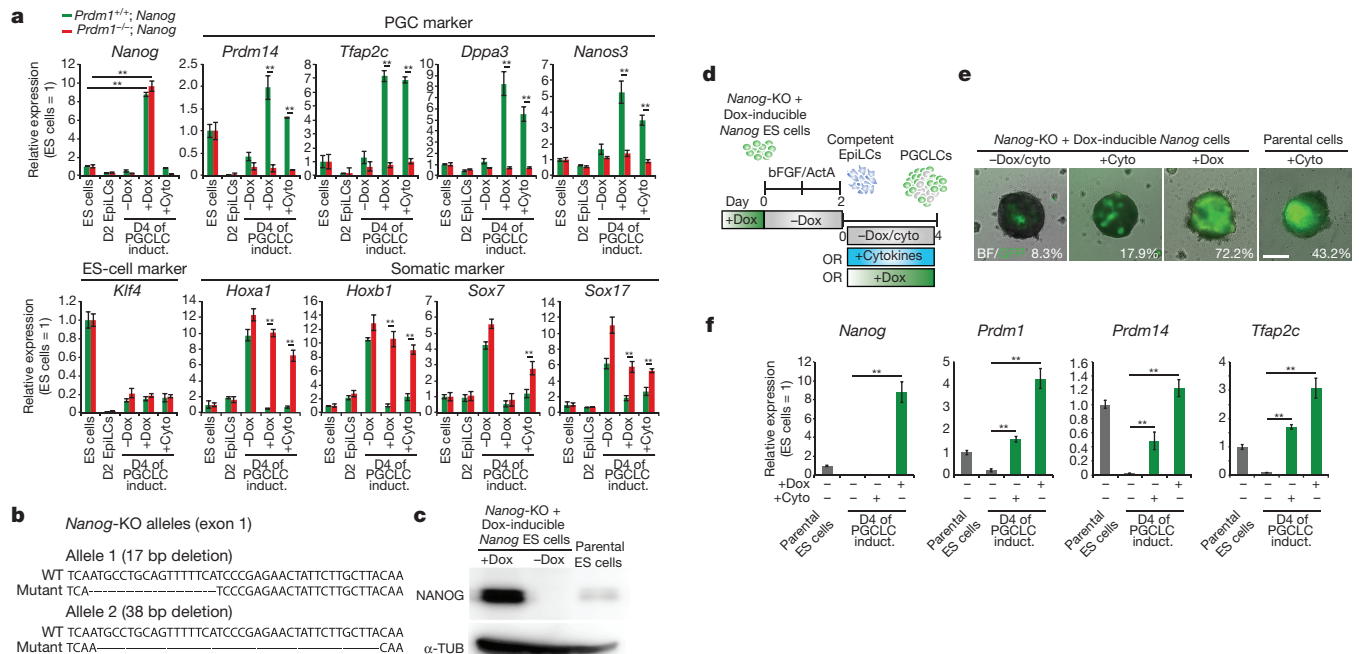


Figure 2 | Loss of *Prdm1* and *Nanog* affects PGCLC specification. **a**, Analysis (qPCR) of mutant (*Prdm1*^{-/-}) and control (*Prdm1*^{+/+}) unsorted cells after *Nanog* expression (+Dox). $\Delta\Delta C_t \pm$ s.d.; $n = 4$ values obtained from two technical replicates from each of two biological replicates. Two-sided/unpaired t -test: ** $P < 0.01$; * $P < 0.05$. cyto, cytokines; D, day; induct., induction. **b**, *Nanog* frameshift mutant alleles. KO, knockout. **c**, Western blot for NANOG and α -tubulin (α -TUB) as depicted. Plus or

minus Dox for 2 days; gel source data are in Supplementary Fig. 1. **d**, Experimental design for **e**, **f**. ActA, activin A. **e**, PGCLC induction in *Nanog*-knockout cells, and rescue by *Nanog* (+Dox). Merged brightfield/GFP images at day 4; GFP⁺ cells (%) after FACS. Scale bar, 200 μ m. **f**, Analysis (qPCR) of ES cells and day-4 PGCLC aggregates shown in **e**. $\Delta\Delta C_t \pm$ s.d.; $n = 4$ values obtained from two technical replicates from each of two biological replicates. Two-sided/unpaired t -test: ** $P < 0.01$.

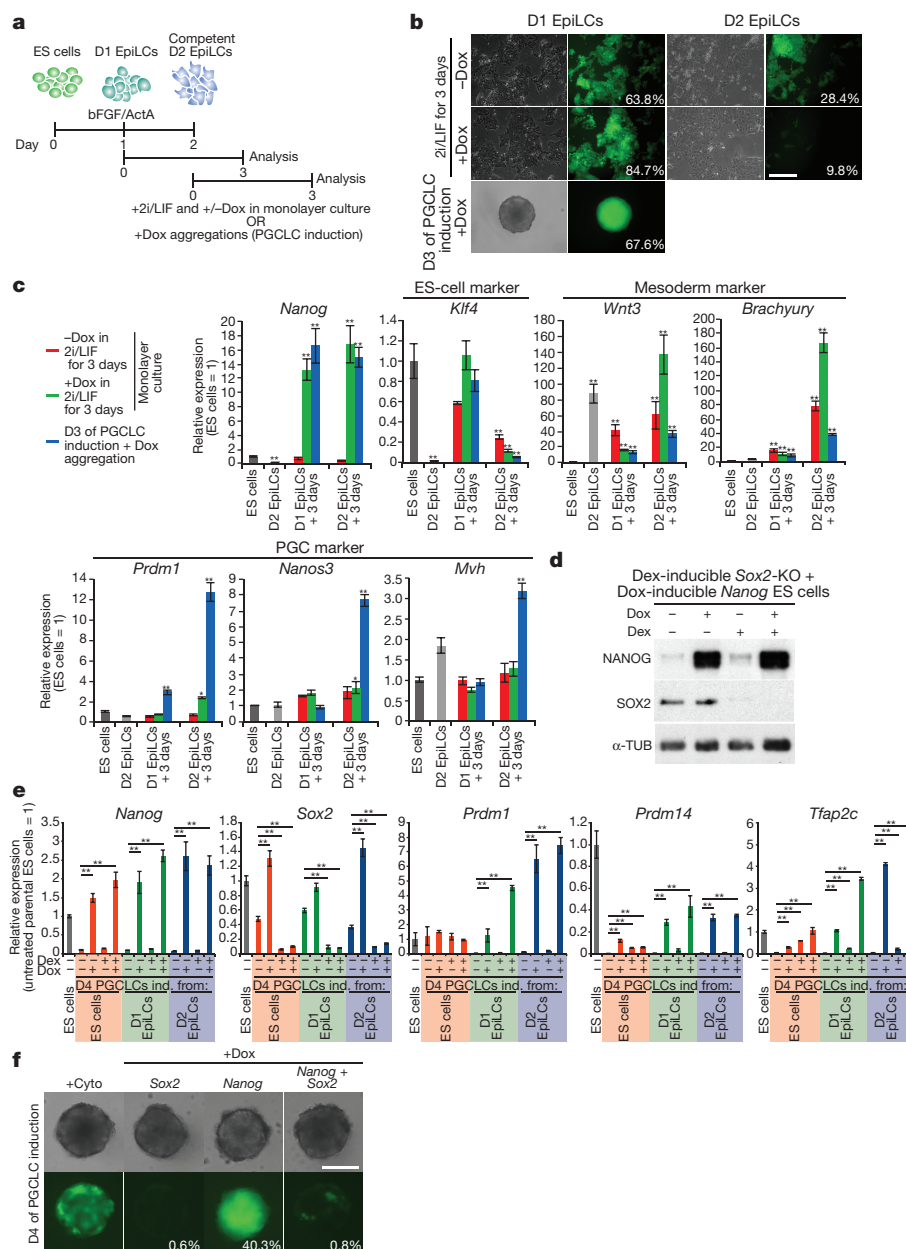


Figure 3 | Competence for PGCLCs versus reversion to ES cells.

a, Experimental design for **b**, **c**. ActA, activin A; D, day. **b**, Brightfield/GFP images depicting day-1/2 EpiLCs in 2i/LIF medium with *Nanog* (+Dox). Note that day-1 EpiLCs revert to ES cells, and attempted PGCLC induction in day-1 EpiLCs also results in ES-cell reversion (see **c**). Scale bar, 200 μ m. **c**, qPCR analysis; $\Delta\Delta C_t \pm$ s.d.; $n = 4$ values obtained from two technical replicates from each of two biological replicates. ES cells were used as a reference for *P* values (two-sided/unpaired *t*-test): ***P* < 0.01; **P* < 0.05. **d**, Western blot for NANOG, SOX2 and α -tubulin (α -TUB) with ES cells

as depicted; plus or minus Dex/Dox for 2 days. Experimental design is in Extended Data Fig. 7b; gel source data are in Supplementary Fig. 1. **e**, Analysis (qPCR) after *Sox2* knockout (KO; +Dex) and *Nanog* induction (ind.; +Dox); $\Delta\Delta C_t \pm$ s.d.; $n = 4$ values obtained from two technical replicates from each of two biological replicates. Parental ES cells were used as a reference for *P* values (two-sided/unpaired *t*-test): ***P* < 0.01; **P* < 0.05. Experimental design is in Extended Data Fig. 7e. **f**, PGCLC induction with Dox-inducible transgenes (*Nanog*, *Sox2* or *Nanog/Sox2*); day-4 brightfield/GFP-GFP⁺ cells (%) after FACS. Scale bar, 200 μ m.

Prdm1^{-/-} embryos *in vivo*²². Furthermore, H3S10ph and γ H2A.X analysis by immunofluorescence of day-6 aggregates indicated that while cell proliferation was unaffected, the rate of apoptosis increased, presumably as the differentiated cells could not survive in the culture conditions (Extended Data Fig. 6c, d).

To investigate PGCLC induction by *Nanog* further, we generated CRISPR-Cas9-mediated *Nanog*-knockout alleles in GOF-GFP ES cells with Dox-inducible *Nanog* (Fig. 2b, c). We found a significant reduction in the induction of PGCLCs from *Nanog*-mutant cells in response to BMP4 (Fig. 2d–f), but ectopic *Nanog* expression rescued this deficit, suggesting complementary roles for BMP4 and *Nanog* in PGCLC induction.

Next, we investigated whether the WNT-BRACHYURY pathway is important for PGCLC induction by *Nanog*, as is the case with BMP4 (ref. 24). We induced PGCLCs in the presence of XAV939 tankyrase inhibitor, which promotes degradation of β -catenin²⁵, resulting in the repression of *Brachyury* (Extended Data Fig. 6e–g). PGCLC induction with BMP4 was repressed by XAV939 but not when induced with *Nanog* (Extended Data Fig. 6h, i). Furthermore, WNT had no detectable effect on *Nanog* expression (Extended Data Fig. 6g, i), indicating that *Nanog* acts independently of WNT-BRACHYURY.

We then asked at which point during the transition of ES cells to EpiLCs cells become responsive to *Nanog* for PGCLC induction. We found that a large majority of day-1 EpiLCs (63.8%) reverted to ES cells

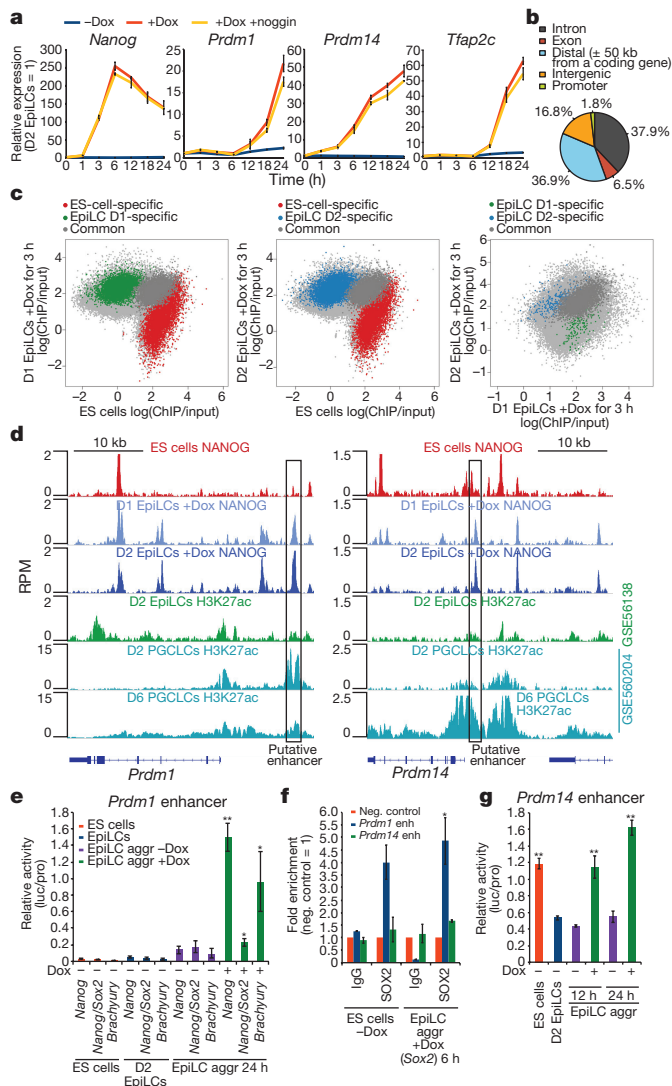


Figure 4 | Context-dependent NANOG binding in ES cells/EpiLCs. **a**, qPCR analysis; induction of *Prdm1*, *Prdm14* and *Tfap2c* by *Nanog* (+Dox) in EpiLCs; $\Delta\Delta C_t \pm$ s.d.; $n = 6$ values obtained from three technical replicates from each of two biological replicates. **b**, Genome-wide NANOG binding in day-2 EpiLCs 3 h after *Nanog* (+Dox). 'Distal' intergenic peaks are ± 50 kb away from coding genes, and those further away are designated as 'intergenic'. **c**, NANOG ChIP-seq in ES cells and day-1/2 EpiLCs, with specific or shared high-confidence peaks; $n = 2$ biological replicates. **d**, day, **d**, ChIP-seq tracks^{9,30} at *Prdm1* and *Prdm14* loci, with putative enhancers (boxed) analysed in **e–g**. RPM, reads per million. **e**, Analysis of *Prdm1* enhancer-luciferase reporter in ES cells, EpiLCs, and after PGCLC induction (+Dox, EpiLC aggregations (aggr), unsorted). Mean luciferase activity normalized to protein quantity (luc/pro) \pm s.d.; $n = 3$ technical replicates. Reference for *P* value (two-sided/unpaired *t*-test): EpiLC aggregations minus Dox; $**P < 0.01$; $*P < 0.05$. Controls and replicates are in Extended Data Fig. 10c, **e**, **f**, SOX2 ChIP-qPCR in ES cells and 6 h after PGCLC induction plus or minus Dox (+/- Sox2) (unsorted EpiLC aggregations). Mean of fold enrichment over negative region \pm standard error of the mean; $n = 4$ values obtained from two technical replicates from each of two biological replicates. Reference for *P* values (two-sided/unpaired *t*-test): IgG; $*P < 0.05$. enh, enhancer; Neg., negative. **g**, Analysis of *Prdm14* enhancer-luciferase reporter in ES cells, EpiLCs and after PGCLC induction (+Dox, EpiLC aggregations, unsorted). Mean luciferase activity normalized to protein quantity \pm s.d.; $n = 3$ technical replicates. Reference for *P* values (two-sided/unpaired *t*-test): EpiLC aggregations minus Dox 24 h; $**P < 0.01$. Colour code is as in **e**; controls and replicates are in Extended Data Fig. 10c, **e**.

when transferred to 2i/LIF medium, and *Nanog* enhanced this response (to 84.7%), as confirmed by expression of *Klf4* and repression of PGC genes (Fig. 3a–c). This reversion to ES cells diminished significantly in day-2 EpiLCs (28.4%), and *Nanog* repressed it further (to 9.8%); instead these cells exhibited a distinct phenotype with expression of *Brachyury* and *Wnt3* mesodermal genes (Fig. 3a–c). Thus, day-2 EpiLCs do not revert to ES cells but acquire competence for PGCLC fate in response to *Nanog*.

Nanog and *Sox2* promote pluripotency in the ICM, but thereafter *Nanog* is detected in the E6.25 posterior epiblast, where PGCs arise^{15,16}, and *Sox2* in the anterior epiblast, where it promotes neuronal fate and inhibits mesodermal specification¹⁶. *Sox2* also represses germline genes in ES cells²⁶ (Extended Data Fig. 7a). We tested the roles of *Nanog* and *Sox2* in our experimental model using ES cells with dexamethasone (Dex)-inducible knockout of *Sox2* (ref. 27), in conjunction with Dox-inducible *Nanog* (Fig. 3d and Extended Data Fig. 7b). Loss of *Sox2* caused a moderate upregulation of *Prdm1* and *Tfap2c* in ES cells without affecting *Nanog* expression (Extended Data Fig. 7c, d). Notably, *Nanog* induced *Prdm1*, *Prdm14* and *Tfap2c* in *Sox2*-knockout day-1 EpiLCs but not in wild-type cells (Fig. 3e and Extended Data Fig. 7e). Since there is a gradual decline in *Sox2* during development of EpiLCs⁵ (Extended Data Fig. 7d), residual *Sox2* in day-1 EpiLCs might repress competency for PGCLCs, and accounts for their reversion to ES cells in response to *Nanog* (Fig. 3a–c). By contrast, Dox-induced expression of *Sox2* strongly repressed PGCLC specification in day-2 EpiLCs in response to *Nanog* but not BMP4 (Fig. 3f and Extended Data Figs 7f–i, 8a, b). *Sox2*, however, caused rapid proliferation of PGCLCs after their induction by BMP4 (Extended Data Fig. 8c), consistent with its role in PGCs *in vivo*²⁸. Thus, progressive downregulation of *Sox2* in EpiLCs contributes to competency for PGCLCs, but thereafter *Sox2* supports proliferation of early germ cells. This further confirms that NANOG and BMP-SMAD act independently during PGCLC induction.

While both *Nanog* and BMP4 induce PGCLCs, the temporal sequence of *Prdm1*, *Prdm14* and *Tfap2c* induction differs slightly. *Nanog* induces *Prdm14* first at 3 h, which increases rapidly over ~ 18 h (Fig. 4a). This is followed by *Prdm1* at ~ 12 h, which increases over the following 12 h, and finally *Tfap2c* expression at ~ 18 h. Thus, all the three regulators of PGCLCs are upregulated within ~ 24 h. While the response of day-2 EpiLCs to BMP4 is similar, *Prdm1* expression is detected first and slightly ahead of *Tfap2c*, followed by *Prdm14* (Extended Data Fig. 9a). *Tfap2c*, a direct target of BLIMP1 (refs 7, 8), is rapidly induced by *Prdm1* alone within 6 h (Extended Data Fig. 9b).

To explore how NANOG promotes both pluripotency and the induction of PGCLCs, we performed NANOG chromatin immunoprecipitation followed by sequencing (ChIP-seq) in ES cells and 3 h after induction of physiological levels (equivalent to ES cells) of NANOG in EpiLCs (Extended Data Fig. 9c). We found NANOG binding primarily in the intergenic regions and introns ($>90\%$), where enhancer elements reside (Fig. 4b and Extended Data Fig. 9d), with marked differences in binding patterns and enriched motifs in ES cells compared to EpiLCs (Fig. 4c and Extended Data Fig. 9e); this provides a basis for the context-dependent functions of NANOG. Overall, many day-2 EpiLC enhancers bound by NANOG show enrichment of H3K27ac in day-2 PGCLCs, indicative of active enhancers²⁹ (Extended Data Fig. 9f–h). This shows that during PGCLC induction, NANOG might contribute to the activation of these elements together with BLIMP1, PRDM14 and AP-2 γ . Importantly, we also found and confirmed intergenic NANOG-binding sites proximate to the *Prdm14* and *Prdm1* loci (Fig. 4d and Extended Data Fig. 10a, b). These sites were devoid of the promoter- and gene-body-associated H3K4me3 and H3K36me3 modifications, respectively. Instead, they were enriched for the enhancer-associated H3K4me1 modification in EpiLCs, suggesting their priming before activation via NANOG and gain of H3K27ac in PGCLCs (Fig. 4d and Extended Data Fig. 10a). Since *Prdm14* is critical for both ES cells and PGCLCs, its enhancer showed a similar H3K4me1/H3K27ac/NANOG enrichment profile in both cell types.

Next, we tested the putative *Prdm1* enhancer in a luciferase reporter assay, and found that following its low activity in ES cells and EpiLCs, *Nanog* activated the enhancer within 24 h after PGCLC induction (Fig. 4e and Extended Data Fig. 10c, d). Notably, *Sox2* strongly repressed this activity, consistent with SOX2 binding to this enhancer (Fig. 4e, f and Extended Data Fig. 10c, d). By contrast, the putative *Prdm14* enhancer, which did not bind SOX2 (Fig. 4f), was active in ES cells; this declined in EpiLCs but increased again within 12 h after the induction of PGCLCs by *Nanog* (Fig. 4g and Extended Data Fig. 10c, e). This reflects the importance of *Prdm14* for both pluripotency and PGCLC fate^{12–14}. Notably, while both BRACHYURY and NANOG bind to and activate the *Prdm1* enhancer²⁴, NANOG acts independently of WNT during PGCLC induction (Fig. 4e and Extended Data Figs 6e–i, 10c, d). Thus, NANOG activates key regulators of PGCLCs independently of BMP4 and WNT signalling. Additional regulatory elements associated with *Prdm1* and *Prdm14* may respond similarly.

The resetting of the epigenome during the gain of competency for PGC-like fate is reflected in the differential NANOG-binding pattern in ES cells and EpiLCs, consistent with the role of NANOG in pluripotency and PGCLC specification (Extended Data Fig. 10f). *Nanog* is detected in the proximal epiblast and the early germ line^{15,16}. Transcription factors also affect competency, since SOX2 inhibits the induction of PGCLCs by NANOG, while NANOG and SOX2 cooperatively promote pluripotency in the ICM/ES cells. NANOG acts independently of BMP4 during PGCLC induction, but they might act cooperatively *in vivo*, since loss of *Nanog* markedly impairs the efficiency of PGCLC specification via BMP4. Notably, epigenome resetting during differentiation of competent EpiLCs establishes a mechanistic paradigm for context-dependent roles of transcription factors such as NANOG that could apply generally during development.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 September 2014; accepted 23 November 2015.

Published online 11 January 2016.

- Yamaguchi, S., Kimura, H., Tada, M., Nakatsuji, N. & Tada, T. *Nanog* expression in mouse germ cell development. *Gene Expr. Patterns* **5**, 639–646 (2005).
- Chambers, I. *et al.* *Nanog* safeguards pluripotency and mediates germline development. *Nature* **450**, 1230–1234 (2007).
- Yamaguchi, S. *et al.* Conditional knockdown of *Nanog* induces apoptotic cell death in mouse migrating primordial germ cells. *Development* **136**, 4011–4020 (2009).
- Carter, A. C., Davis-Dusenbery, B. N., Koszka, K., Ichida, J. K. & Eggan, K. *Nanog*-independent reprogramming to iPSCs with canonical factors. *Stem Cell Reports* **2**, 119–126 (2014).
- Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532 (2011).
- Lawson, K. A. *et al.* *Bmp4* is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev.* **13**, 424–436 (1999).
- Magnúsdóttir, E. *et al.* A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nature Cell Biol.* **15**, 905–915 (2013).
- Nakaki, F. *et al.* Induction of mouse germ-cell fate by transcription factors *in vitro*. *Nature* **501**, 222–226 (2013).
- Buecker, C. *et al.* Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 838–853 (2014).
- Zylicz, J. J. *et al.* Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. *eLife* **4**, e09571 (2015).
- Surani, M. A., Hayashi, K. & Hajkova, P. Genetic and epigenetic regulators of pluripotency. *Cell* **128**, 747–762 (2007).
- Ma, Z., Swigut, T., Valouev, A., Rada-Iglesias, A. & Wysocka, J. Sequence-specific regulator *Prdm14* safeguards mouse ESCs from entering extraembryonic endoderm fates. *Nature Struct. Mol. Biol.* **18**, 120–127 (2011).

- Yamaji, M. *et al.* Critical function of *Prdm14* for the establishment of the germ cell lineage in mice. *Nature Genet.* **40**, 1016–1022 (2008).
- Grabole, N. *et al.* *Prdm14* promotes germline fate and naive pluripotency by repressing FGF signalling and DNA methylation. *EMBO Rep.* **14**, 629–637 (2013).
- Sun, L. T. *et al.* *Nanog* co-regulated by Nodal/Smad2 and Oct4 is required for pluripotency in developing mouse epiblast. *Dev. Biol.* **392**, 182–192 (2014).
- Hoffman, J. A., Wu, C. I. & Merrill, B. J. Tcf7l1 prepares epiblast cells in the gastrulating mouse embryo for lineage specification. *Development* **140**, 1665–1675 (2013).
- Seki, Y. *et al.* Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev. Biol.* **278**, 440–458 (2005).
- Hajkova, P. *et al.* Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452**, 877–881 (2008).
- Hackett, J. A. *et al.* Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* **339**, 448–452 (2013).
- Koshimizu, U., Watanabe, M. & Nakatsuji, N. Retinoic acid is a potent growth activator of mouse primordial germ cells *in vitro*. *Dev. Biol.* **168**, 683–685 (1995).
- West, J. A. *et al.* A role for Lin28 in primordial germ-cell development and germ-cell malignancy. *Nature* **460**, 909–913 (2009).
- Ohinata, P. *et al.* *Blimp1* is a critical determinant of the germ cell lineage in mice. *Nature* **436**, 207–213 (2005).
- Bao, S. *et al.* The germ cell determinant *Blimp1* is not required for derivation of pluripotent stem cells. *Cell Stem Cell* **11**, 110–117 (2012).
- Aramaki, S. *et al.* A mesodermal factor, *T*, specifies mouse germ cell fate by directly activating germline determinants. *Dev. Cell* **27**, 516–529 (2013).
- Huang, S.-M. A. *et al.* Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. *Nature* **461**, 614–620 (2009).
- Adachi, K. *et al.* Context-dependent wiring of *Sox2* regulatory networks for self-renewal of embryonic and trophoblast stem cells. *Mol. Cell* **52**, 380–392 (2013).
- Masui, S. *et al.* Pluripotency governed by *Sox2* via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature Cell Biol.* **9**, 625–635 (2007).
- Campolo, F. *et al.* Essential role of *Sox2* for the establishment and maintenance of the germ cell line. *Stem Cells* **31**, 1408–1421 (2013).
- Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA* **107**, 21931–21936 (2010).
- Kurimoto, K. *et al.* Quantitative dynamics of chromatin remodeling during germ cell specification from mouse embryonic stem cells. *Cell Stem Cell* **16**, 517–532 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank H. Leitch for ES cell lines, C. Lee for help with animal husbandry, H. Niwa for vectors and conditional *Sox2*-knockout ES cells, N. Miller, R. Walker and A. Riddell for FACS sorting and J. Bauer for analysis of microarray data. K.M. was supported by the Japan Society for the Promotion of Science (JSPS) Institutional Program for Young Researchers Overseas Visits. U.G. was supported by a Marie Skłodowska-Curie and a Newton Trust/Leverhulme Trust Early Career fellowship. J.J.Z. was a recipient of a Wellcome Trust PhD Studentship (RG44593). T.K. was supported by a JSPS Fellowship for research abroad. This research was supported by Gurdon Institute core grants from the Wellcome Trust (092096) and Cancer Research UK (C6946/A14492), and a grant from the Wellcome Trust to M.A.S. (WT096738).

Author Contributions K.M. and U.G. designed and performed experiments, and wrote the paper; W.W.C.T. designed and carried out the luciferase assays; NANOG ChIP experiments were carried out by J.J.Z., while R.S. performed immunofluorescence analysis; T.K. and S.K. designed and carried out the chimera experiments; S.D. performed bioinformatic analysis; R.B. developed the 'Object Scan' plugin; M.A.S. supervised the project, designed experiments and wrote the paper. All authors discussed the results and contributed to the manuscript.

Author Information Microarray data have been deposited in the Gene Expression Omnibus under accession number GSE71933. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.S. (a.surani@gurdon.cam.ac.uk).

METHODS

Animals. Timed natural matings were used for all experiments, where noon of the next day after the vaginal plugs of mated females were identified was scored as E0.5. Animal studies were authorized by a UK Home Office Project License and carried out in a Home Office-designated facility. No statistical methods were used to predetermine sample size.

Mouse embryonic stem cells. $\Delta PE-Oct3/4$ -GFP (GOF-GFP)^{31,32}, *Prdm1*-GFP and *Prdm1*^{-/-} ES cell lines were established previously^{14,19,22,23}. Inducible *Sox2*-knockout (2CG2) ES cell line was a gift from H. Niwa²⁷. All ES cell lines were maintained in naive 'ground state'³³ condition; that is, in N2B27 medium (R&D) with 2i (PD0325901, 1 μ M; CHIR99021, 3 μ M; Stemgent), LIF and 1% KnockOut Serum Replacement (KSR; Life Technologies) on fibronectin-coated dishes (16.7 μ g ml⁻¹; Millipore). Medium was changed daily. ES cell colonies were passaged by dissociating with TrypLE (Life Technologies).

Establishment of PiggyBac-based Tet-on expression system during PGCLC induction. *Oct3/4*, *Sox2*, *Nanog*, *Prdm1*, *Prdm14* and *Brachyury* complementary DNAs were cloned from mouse cDNA pool. cDNAs were inserted into PiggyBac-based doxycycline (Dox)-inducible vectors (a gift from H. Niwa). These vectors were transfected using Lipofectamine 2000 (Life Technologies) into ES cells together with a pPyCAG-PBase vector and a pPBCAG-rtTAIRESNeo^r vector, which harbours a neomycin resistance gene. After 1 week of neomycin (80 μ g ml⁻¹; Life Technologies) selection, pooled or single clones were used for experiments. To induce transgene expression, various concentrations of Dox (Sigma-Aldrich) were added to the media.

Induction of EpiLCs and PGCLCs. EpiLCs and PGCLCs were induced as described previously⁵. Transgenes were induced by addition of Dox at day 0 of PGCLC induction. 100 ng ml⁻¹ Dox was used in experiments shown in Figs 3f, 4e, g and Extended Data Figs 3b, c, 7g–i, 8a–c. 200 ng ml⁻¹ Dox was used in Fig. 4b–d, f. 700 ng ml⁻¹ of Dox was used in all other experiments. PGCLCs were induced as described in the manuscript. For inhibition of the BMP-SMAD pathway, noggin (200 ng ml⁻¹; R&D) was added to the media at day 0 of PGCLC induction. For inhibition of WNT signalling, XAV939 (1 μ M; Sigma-Aldrich) was added to the media.

Reversion of epiblast-like cells into ES-like cells. Day-1 or day-2 EpiLCs were transferred into GMEM 15%KSR 2i/LIF with or without Dox in monolayer culture. In addition, day-1 or day-2 EpiLCs were aggregated in low-cell-binding U-bottom-shaped 96-well plates (Thermo Scientific) (1,000–2,000 cells per well) in PGCLC induction media (GMEM with L-glutamine (Life Technologies), 15% KSR (Life Technologies), 1 \times MEM NEAA (Life Technologies), 1 \times sodium pyruvate (Life Technologies), 1 \times 2-mercaptoethanol (Life Technologies), 1 \times penicillin/streptomycin (Life Technologies)) and Dox. The medium was replaced daily. After 3 days, the GFP reporter signal was analysed with a fluorescence microscope and via FACS analysis. RNA was collected from pooled cells for qRT-PCR.

EGCLC derivation. Day-4 aggregates were dissociated with TrypLE and plated on mitomycin C-treated mouse embryonic fibroblast (MEF) feeder cells with PGC selection medium (DMEM with L-glutamine (Life Technologies), 15% fetal bovine serum (FBS; Sigma-Aldrich)), LIF, 15 ng ml⁻¹ bFGF, 30 ng ml⁻¹ SCF (R&D) and 2 μ M all-trans-retinoic acid (Sigma-Aldrich). Retinoic acid promotes germ cell self-renewal while promoting differentiation of ES cells^{20,21}. The media was replaced daily. After 5 days, proliferating GFP⁺ cells were dissociated with TrypLE and plated on fibronectin-coated dishes with ESC medium (N2B27 with 2i/LIF).

FACS. PGCLCs were dissociated with TrypLE, washed with DMEM containing 10% FBS and resuspended with 1 \times PBS containing 0.1% BSA. Large clumps of cells were removed using a cell strainer (BD Biosciences). The cells were analysed and sorted on flow cytometers (FACS Calibur, BD Biosciences; MoFlo high speed cell sorter, Beckman Coulter; S3 cell sorter, Biorad).

RT-qPCR. Total RNAs from ES cells, EpiLCs and FACS-sorted cells were extracted using the RNeasy Mini Kit (Qiagen) or Picopure RNA Isolation Kit (Life Technologies). The total RNAs were reverse transcribed by the Quantitect Reverse Transcription Kit (QIAGEN). The first-strand cDNAs were used for RT-qPCR analysis with SYBR Green PCR reagent (Sigma-Aldrich). The primer sequences used for the qRT-PCR are listed in Supplementary Table 1. Student's *t*-test was used to test for significance.

Microarray. ES cells and day-4 PGCLCs were dissociated and sorted with a MoFlo high-speed cell sorter (Beckman Coulter). Total RNAs were extracted using the RNeasy Mini Kit (QIAGEN). Complementary RNA (cRNA) generation, quality control, hybridization and data analysis were performed by Cambridge Genomic Services at the University of Cambridge. Raw intensity values from Illumina MouseWG-6 v.2.0 expression beadchip microarrays were pre-processed with the Bioconductor lumi and preprocessCore packages (<http://www.bioconductor.org>): Probes that were not detected in at least one sample were removed, Variance stabilization transformation (VST) was applied, and samples were quantile-normalized. Differential expression was evaluated with the Bioconductor limma package.

Comparison with published microarray data (Extended Data Fig. 3j). Our data set was assayed on an Illumina MouseWG-6 v.2.0 expression beadchip, the data set from ref. 5 was assayed on the Affymetrix Mouse Genome 430 2.0 Array platform. We therefore quantile-normalized the data sets to ensure that the data sets span comparable ranges of expression values. PCA was performed on the centre-scaled expression values, where systematic differences between platforms are mainly captured by the first principal component.

Immunofluorescence stainings. Day-3, day-4 and day-6 aggregates were fixed with 2% or 4% paraformaldehyde for 20 min at room temperature or for 2 h at 4°C. Fixed aggregates were washed several times in PBS and transferred into 10% sucrose/PBS (2 h), 20% sucrose/PBS (2 h) and finally into OCT embedding matrix (overnight; CellPath). Next day, cell aggregates were embedded in OCT in tissue moulds and stored at -80°C. A Leica Cryostat CM3050S was used to cut the OCT blocks in 6–8- μ m-thick sections, which were collected on SuperFrost Plus slides (VWR).

For immunofluorescence staining, the slides were washed with PBS, permeabilized with PBS/0.1–1% Triton X-100 and then incubated with primary antibodies in permeabilization buffer including 5% donkey serum (Sigma-Aldrich) overnight at 4°C. Next day, slides were washed with PBS and incubated with secondary antibodies in permeabilization buffer for 2 h at room temperature, washed with PBS, incubated with 4',6-diamidino-2-phenylindole (DAPI) in PBS for 15–30 min, and mounted using Vectashield Mounting Medium (VECTOR Labs). Images were acquired using a Leica SP5 or SP8 confocal microscope. For 5hmC stainings, it was required to perform an additional antigen retrieval step before incubation with primary antibodies: slides with sections were transferred into TE buffer, pH 8, at ~95°C and microwaved at very low power for 45 min.

The following primary antibodies were used: mouse anti-*OCT3/4* (1:100, BD Biosciences, O50808), rat anti-BLIMP1 (1:50, eBioscience, clone 6D3, 14-5963), rabbit anti-AP-2 γ (1:250, SantaCruz, sc-8977), rabbit anti-PRDM14 (1:250, a gift from D. Reinberg), rabbit anti-DAZL (1:500, Abcam, ab34139), mouse anti-H3K9me2 (1:250, Abcam, ab1220 and 1:500, Millipore, 07-441), rabbit anti-H3K27me3 (1:500, Millipore, 07-449), rabbit anti-TET1 (1:500, Millipore, 09-872), rabbit anti-5hmC (1:500, Active Motif, 39791), goat anti-KLF4 (1:100, R&D, AF3158), rabbit anti-H3S10ph (1:500, Millipore, 06-5770), mouse anti- γ H2A.X (1:250, Millipore, 05-636), rat anti-GFP (1:500, Nakalai Tesque, GF090R). Alexa Fluor488 and 568 were used as secondary antibodies (1:500, Life Technologies).

Quantification of immunofluorescence data. All quantifications were performed using Fiji³⁴. The DAPI, H3S10ph and γ H2A.X channels were processed by applying a Gaussian Blur (H3S10ph staining: DAPI/H3S10ph: σ 0.5/1.1; DAPI/ γ H2A.X: σ 1.0/1.5) to reduce noise. The images were then binarized using the Otsu thresholding algorithm and holes were filled before the total signal area was measured. In day-6 *Prdm1*^{-/-} plus Dox aggregates, many cells underwent cell death. Therefore, nuclei with bright discrete spots of DAPI signal, which indicates chromatin condensation, were excluded from the analysis. The diameter of ~10 cells was measured and used to calculate the average area of one cell to estimate the number of cells in the field of view (DAPI⁺ area/area of one cell).

For all other quantifications on a single-cell level, we developed 'Object Scan', which is an object mapping and analysis plugin for Fiji that combines advanced functions with a user-friendly interface. Images are processed with a choice of feature enhancement algorithms, objects are identified by patch sampling to detect intensity edges based on the local energy gradient, and the generated two-dimensional masks are clustered in three dimensions to define the final object map for analysis. We used Object Scan to carry out DoG processing and contained signal analysis using the DAPI channel for object mapping, watershed segmentation, a scan radius of one and the following channel specific settings: edge gradient = 10, estimated object radius = 9 μ m. The results were scale normalized ($(X - X_{\min})/(X_{\max} - X_{\min})$) to the range 0 to 1 for comparison. Student's *t*-test was used to test for significance. The Object Scan plugin is available from this link: <http://www.gurdon.cam.ac.uk/stafflinks/downloadspublic/imaging-plugins>.

ChIP. Low cell number ChIP-qPCR was performed as previously described³⁵. 3×10^5 cells per ChIP were fixed in 1% formaldehyde (room temperature, 10 min), quenched with 1 vol. of 250 mM glycine (room temperature, 5 min), and rinsed with chilled TBSE buffer (20 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA) twice before storage at -80°C. After thawing the cells on ice, fixed cells were lysed with 100 μ l 1% SDS lysis buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 1% SDS, Roche protease inhibitor cocktail; on ice, 5 min) and then centrifuged (2,000 r.p.m., 10 min). Pellet was resuspended in 100 μ l of dilution buffer (16.7 mM Tris-HCl, pH 8, 167 mM NaCl, 1.2 mM EDTA, 1.1% Triton X-100, 0.01% SDS, Roche protease inhibitor cocktail). Samples were sonicated nine times (30-s pulses with 30-s break interval) using the Bioruptor water bath sonicator (Diagenode). Chromatin extracts were then pre-cleared with Dynal Magnetic Beads (Invitrogen) (4°C, 1 h) followed by centrifugation (2,000 r.p.m., 30 min). Supernatant (pre-cleared chromatin) was immunoprecipitated overnight with Dynal Magnetic Beads

coupled with anti-NANOG antibody (1 µg per ChIP, Cosmo Bio Co., RCAB0001P) or normal rabbit serum (1 µg per ChIP). On the next day, beads were washed (nutate in wash buffer for 5 min at 4 °C) in low-salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.0, 150 mM NaCl), high-salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.0, 300 mM NaCl) and LiCl buffer (0.25 M LiCl, 1% NP400, 1% Na deoxycholate, 1 mM EDTA, 10 mM Tris-HCl, pH 8.0), for a total of three washes. Following an additional wash in TE, elution was performed in a PCR machine (68 °C, 10 min). After digesting and reverse crosslinking (with Proteinase K at 42 °C for 2 h and 68 °C for 6 h) DNA was purified (phenol-chloroform extraction) and used for qPCR analysis. For the negative control region, we used the *Snai3* locus as described previously³⁶. Student's *t*-test was used to test for significance.

The same protocol was used for the SOX2 ChIP with some deviations. Day-2 EpiLCs were aggregated in low-binding plates for 6 h in the presence of 200 ng ml⁻¹ of Dox before collection. 5 × 10⁶ ES cells and EpiLCs, respectively, were fixed and processed as described earlier. Samples were sonicated 20 times (30-s pulses with 30-s break interval) using a Bioruptor water bath sonicator (Diagenode). Samples were divided for immunoprecipitations with SOX2 antibody (10 µg per ChIP, Santa Cruz, sc-17320 X) or normal rabbit IgG (10 µg per ChIP, Santa Cruz, sc-2027 X) as a negative control. Beads were washed with low-salt buffer, and twice with high-salt buffer for 10 min each. The beads were rinsed in TE, resuspended in Proteinase K digestion buffer (20 mM HEPES, 1 mM EDTA, 0.5% SDS) with 2 µl of 10 mg ml⁻¹ Proteinase K and incubated for 15 min at 50 °C. In parallel, 2 µl of 10 mg ml⁻¹ Proteinase K was added to the saved input samples. Three microlitres 5 M NaCl was added to the supernatants and the input samples. To reverse the crosslinks, samples were incubated at 42 °C for 2 h and 68 °C overnight. Next day, the DNA was purified using Agencourt Ampure XP beads (Beckman Coulter) according to the manufacturer's instructions. The purified DNA was used for qPCR analysis. For the negative control region, we used the *Snai3* locus. Student's *t*-test was used to test for significance. The primer sequences used for RT-qPCRs are listed in Supplementary Table 1.

NANOG ChIP-seq. The NANOG ChIP for subsequent sequencing was performed as described earlier with some deviations. Day-1 or day-2 EpiLCs were aggregated in low-binding plates for 3 h in the presence of 200 ng ml⁻¹ of Dox. ES cells and EpiLCs were fixed and processed as described earlier. 3 × 10⁶ fixed cells were lysed with 1 ml 1% SDS lysis buffer and then centrifuged (2,000 r.p.m., 15 min). Nuclear fraction was resuspended in 0.9 ml of dilution buffer. Samples were sonicated ten times (30-s pulses with 30-s break interval) using a Bioruptor water bath sonicator (Diagenode). Immunoprecipitations were performed with anti-NANOG antibody (2 µg per ChIP, Cosmo Bio Co., RCAB0001P). After elution, samples were digested with Proteinase K and reverse crosslinked for 6 h at 68 °C. Twelve nanograms of purified DNA was used for library preparation using Ovation Ultralow DR Multiplex System (Nugen). Once prepared, library was size selected and sequenced using HiSeq2000 with single-end 50 nucleotides read length.

ChIP-seq analysis. ChIP-seq reads were aligned with the bwa aligner (<http://bio-bwa.sourceforge.net>) to the mouse reference genome (GRCm38/mm10). Peaks were called with MACS (version 2.1.0; <https://github.com/taoliu/MACS>) and visualized using the Integrative Genomics Viewer (<https://www.broadinstitute.org/igv/>). Peak regions from two biological replicates were intersected using bedops (<http://bedops.readthedocs.org>). Overlapping peak regions with peak summits within <50 nucleotides distance in both replicates were retained. Peak regions from the three cell types were merged. Differences in ChIP-seq read intensities on peak regions were evaluated by using diffReps (<http://code.google.com/p/diffreps>) and MACS (macs2 bdgdiff). High-confidence sets of differentially bound regions that were detected by both methods were selected for further analysis by applying the following thresholds for diffReps: pValue < 0.001 and abs(log2FC) > 1. Previously published H3K27ac ChIP-seq data sets^{9,30} were aligned to the mouse reference genome in a similar manner as described earlier, and H3K27ac enrichment (log(ChIP/input) values were determined on NANOG peak regions.

De novo motif analysis. High-confidence MACS peaks, for which the distance of the peak summits in both replicates was <50 nucleotides, were selected. *De novo* motifs were determined with HOMER (<http://homer.salk.edu/homer>) in the 2,000 top-enriched peaks in ES cells, day-1 and day-2 EpiLCs for both repeat-masked and repeat-unmasked regions within ± 50 nucleotides of the peak summit.

Luciferase assay. Genomic regions containing putative enhancers of *Prdm1* and *Prdm14*, as well as a negative control region depleted of enhancer signatures, were amplified from mouse E14 ES-cell genomic DNA. These regions were cloned into a PiggyBAC-based firefly luciferase reporter plasmid upstream of a minimal TK promoter. Stable luciferase reporter GOF-GFP ES cell lines, which can overexpress *Nanog*, *Nanog/Sox2* or *Brachyury* upon Dox addition, were established. Cell pellets were collected from ES cells cultured in N2B27 2i/LIF, day-2 EpiLCs and

EpiLCs after PGCLC induction ± Dox at 12/24 h. Luciferase assays were performed with the ONE-Glo Luciferase Assay System (Promega). Protein concentration in each lysate was quantified by Pierce 660 nm Protein Assay (Thermo Scientific). Relative luciferase activities were obtained by dividing luciferase activity by protein concentration in each sample.

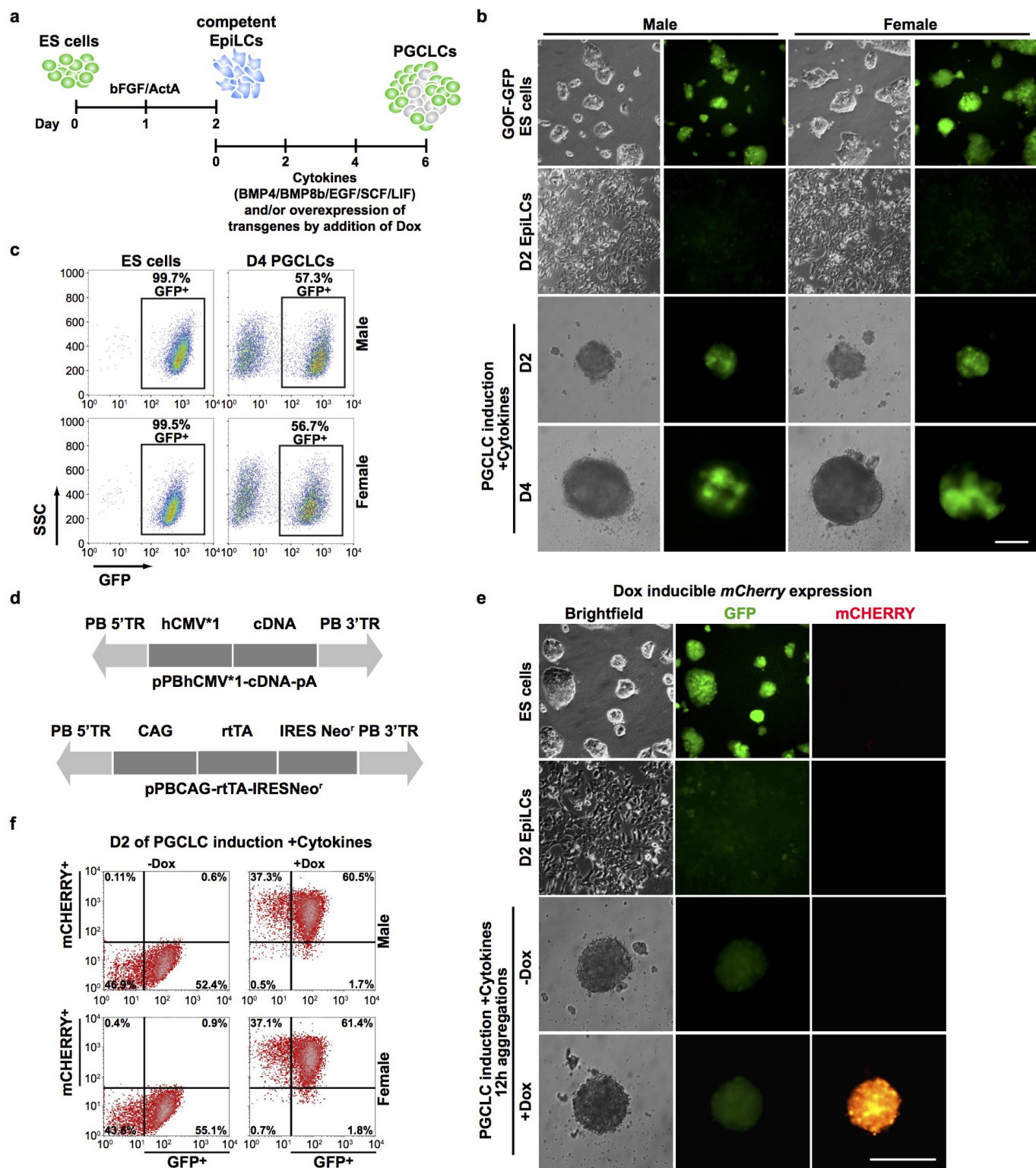
Blastocyst injections. ES-cell clones carrying both the *Nanog* transgene and a CAG monomeric Kusabira-orange (mKO) fluorescence reporter were selected by neomycin (Sigma-Aldrich) and zeocine (Life Technologies). Day-4 PGCLCs were induced from day-2 EpiLCs with *Nanog* and used for derivation of EGCLCs. For ES cells or day-4 PGCLC injections, GOF-GFP ES cells were co-transfected with a vector, which enabled inducible expression of *Nanog* and constitutive expression of Venus, a variant of eGFP. For day-4 PGCLCs, after induction of PGCLCs with *Nanog*, cells were stained with PE-conjugated-CD61 antibody (1:10, Biolegend, 104308) and Alexa660-conjugated-SSEA-1 antibody (2.5 µl per 105 cells, eBioscience, clone eBioMC-480, 50-8813) according to the manufacturer's instructions. Double-positive PGCLC cells were collected by using a S3 cell sorter (Biorad). Embryos for chimaera experiments were obtained from CBA/C57BL/6 F1 crossed with C57BL/6 mice. Blind tests or randomization methods were not used. The sex of embryos was not determined. Manipulations of embryos were performed as described previously³⁷. Briefly, five cells were injected into a morula, which were subsequently cultured in KSOM (Millipore). On the following day, the embryos were transferred into the uteri of pseudopregnant mice. All embryos were analysed 1 week after embryo transfer, which corresponded to E9.5.

Generation of *Nanog*-knockout ES cells. The CRISPR-Cas9 system was used to generate *Nanog*-knockout ES cells. Guide RNAs (gRNAs) targeting exon 1 of the *Nanog* gene were cloned into pX330 (Addgene)³⁸. One microgram of this plasmid was transfected with a pPyCAG-monomeric Kusabira Orange-IRES-Pac plasmid. Transfected cells were selected by puromycin (1 µg ml⁻¹) for 2 days. Clonal *Nanog*-knockout ES cell lines were established and mutations of *Nanog* alleles were confirmed by qPCR, western blotting and DNA sequencing. Subsequently, pPBh-CMV*1-*Nanog*-pA plasmid was transfected into those lines with pPyCAG-PBase and pPBhCMV*1-rtTA-IRESNeo^r to generate *Nanog*-knockout ES cell lines carrying a Dox-inducible *Nanog* transgene. Loss of *Nanog* affected the growth of ES cells. Thus, these cell lines were maintained in N2B27 2i/LIF with a low dose of Dox (100 ng ml⁻¹). gRNA sequences are listed in Supplementary Table 1.

Western blots. 5 × 10⁴ cells were lysed in lysis buffer (50 mM Tris-HCl (pH 8.0), 1% SDS, 10 mM EDTA). Protein concentration was measured by Bicinchoninic Acid Kit (Sigma-Aldrich). The protein amount was adjusted among samples, then 4 × Laemmli buffer was added. Samples were boiled at 95 °C for 5 min. Proteins were separated on 10% acrylamide gels, blotted on Immobilon-P transfer membrane (Millipore). The membrane was blocked with 5% skimmed milk and incubated with primary antibodies: anti-NANOG (1:500, mouse IgG, eBioscience, clone eBioMLC-51, 14-5761), anti-SOX2 (1:500, rabbit IgG, Cell Signaling, 2748), anti-α-tubulin (1:1,000, mouse IgG, Sigma-Aldrich, clone DM1A, T9026). Primary antibodies were detected on X-ray film with anti-rabbit or anti-mouse IgG conjugated with horseradish peroxidase (Dako) followed by detection using Western Detection System (GE Healthcare). For gel source data, see Supplementary Fig. 1.

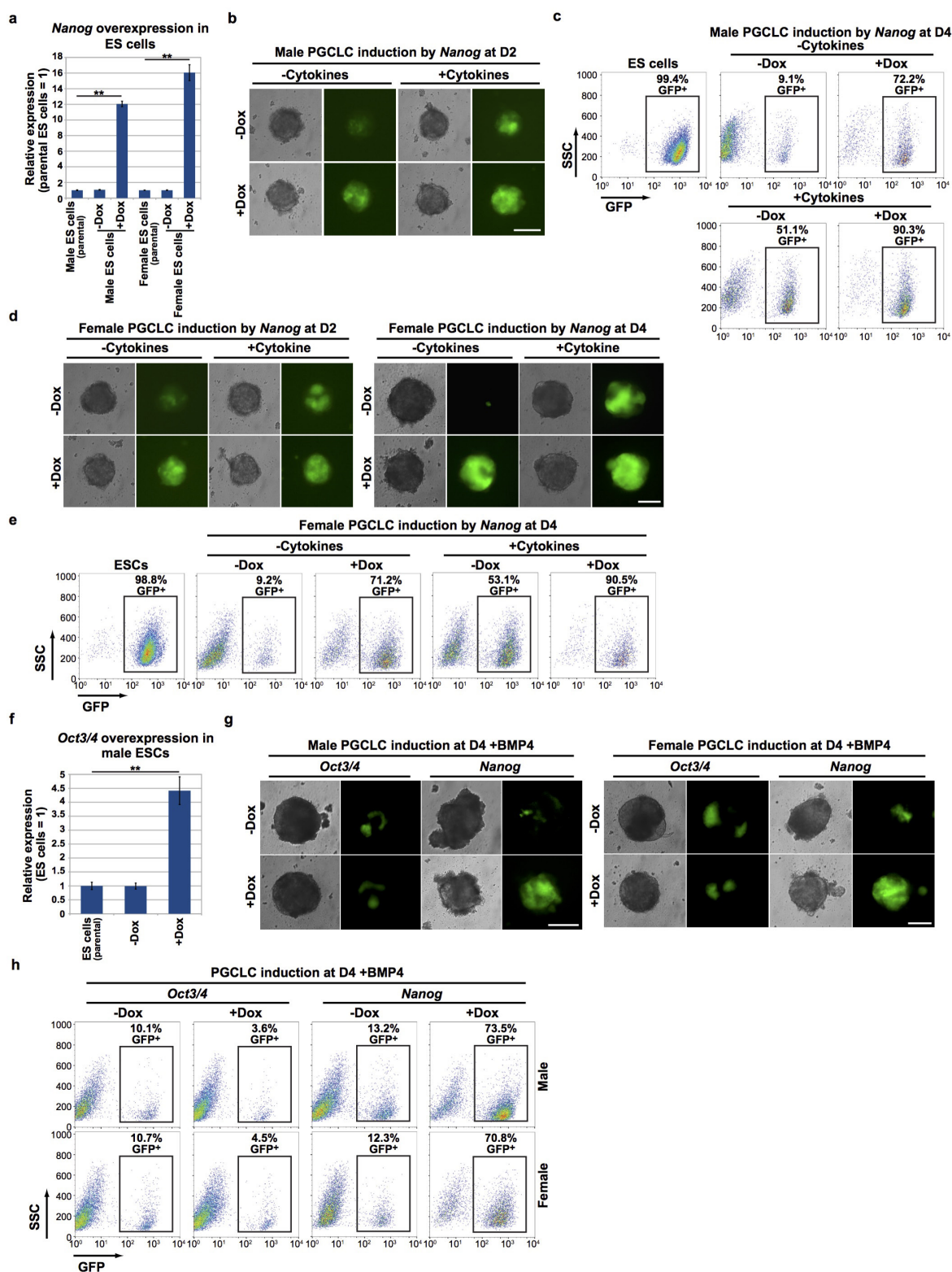
Generation of *Sox2*-conditional-knockout ES cells with Dox-inducible *Nanog* transgene. pPBhCMV*1-*Nanog*-pA, pPBhCMV*1-rtTA-IRESNeo^r and pPyCAG-PBase were transfected into the *Sox2*-conditional-knockout ES cell line (2CG2)²⁷. After 1 week of neomycin selection (80 µg ml⁻¹), pooled cells were used for the subsequent experiments. Dexamethasone-inducible *Sox2*-knockout and Dox-inducible *Nanog* expression were confirmed by qPCR and western blotting.

- Yeom, Y. I. *et al.* Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonic cells. *Development* **122**, 881–894 (1996).
- Yoshimizu, T. *et al.* Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. *Dev. Growth Differ.* **41**, 675–684 (1999).
- Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **433**, 519–523 (2008).
- Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676–682 (2012).
- Ng, J.-H. *et al.* In vivo epigenomic profiling of germ cells reveals germ cell molecular signatures. *Dev. Cell* **24**, 324–333 (2013).
- Gillich, A. *et al.* Epiblast stem cell-based system reveals reprogramming synergy of germline factors. *Cell Stem Cell* **10**, 425–439 (2012).
- Behringer, R., Gertsenstein, M., Nagy, K. V. & Nagy, A. *Manipulating the Mouse Embryo. A Laboratory Manual* 3rd edn, 453–506 (2003).
- Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
- Marks, H. *et al.* The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590–604 (2012).



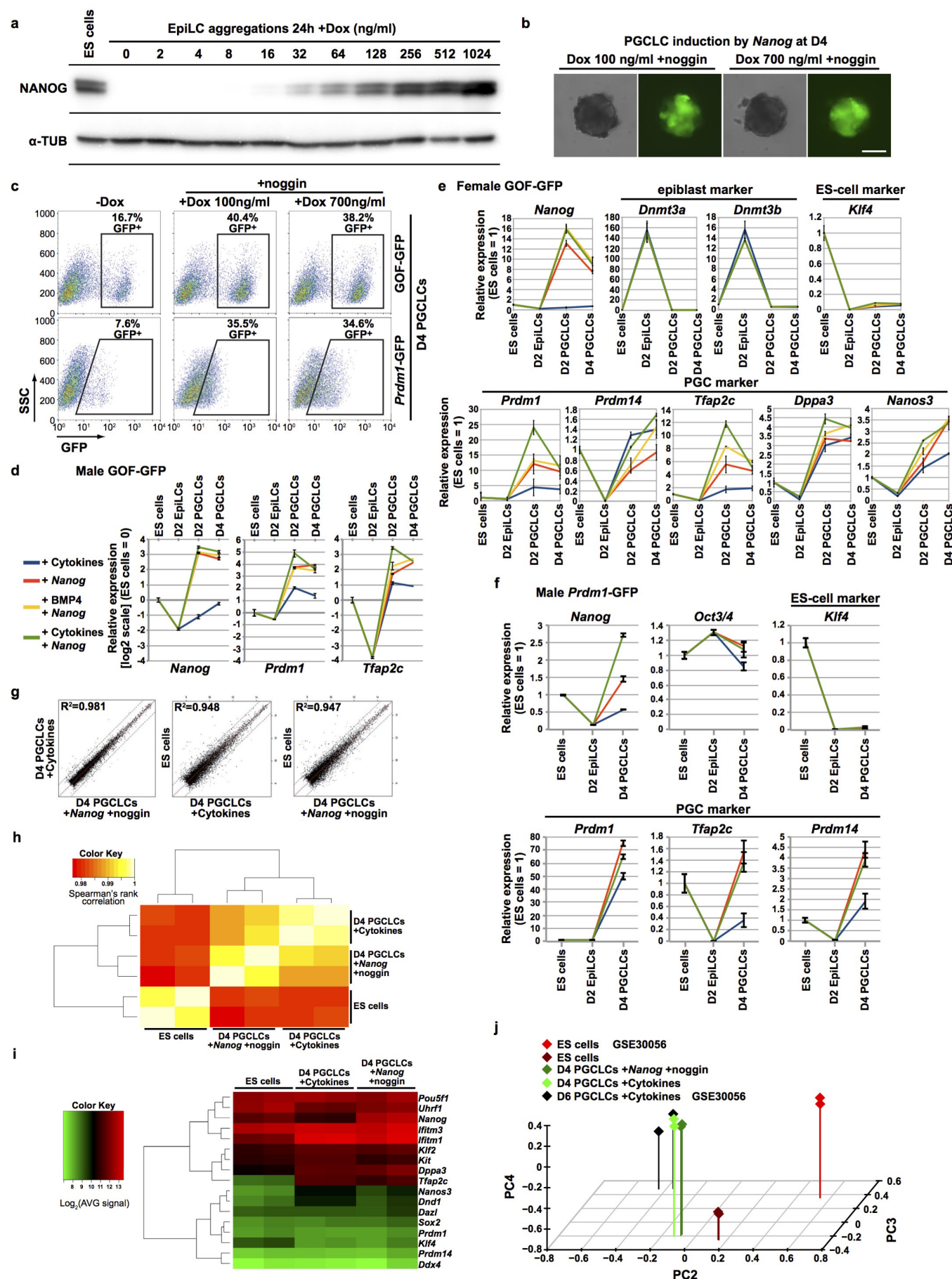
Extended Data Figure 1 | GOF-GFP as a reporter for PGCLCs; Dox-inducible transgene system. **a**, Experimental design for the induction of PGCLCs *in vitro*. **b**, Representative brightfield/GFP images of male/female GOF-GFP ES cells, day-2 EpiLCs, day-2 and day-4 cytokine-induced PGCLCs. Scale bar, 200 μ m. **d**, day. **c**, FACS analysis for GFP with samples shown in **b**. SSC, side scatter. **d**, Simplified scheme of PiggyBac (PB 5'TR and PB 3'TR) based plasmids for transgene overexpression using the Tet-On system. The rtTA protein activates the minimal promoter (hCMV*-1) driving the expression of the cDNA of interest only in the presence of doxycycline (Dox). **e**, Proof-of-principle experiment to

test the Dox-inducible expression of a transgene during the sequential differentiation of PGCLCs from ES cells. GOF-GFP ES cells carrying PiggyBac based Dox-inducible *mCherry* expression plasmids were differentiated into day-2 EpiLCs and then induced into PGCLCs with cytokines in plus or minus Dox conditions. Representative brightfield, GFP and mCherry images are shown 12 h after aggregation. Scale bar, 200 μ m. **f**, Representative FACS analysis for GFP and mCherry of day-2 cytokine-induced PGCLCs from male/female GOF-GFP ES cells carrying a Dox-inducible *mCherry* transgene. Most cells express mCherry after Dox addition.



Extended Data Figure 2 | *Nanog* but not *Oct3/4* induces GFP⁺ cells from competent EpiLCs. **a**, qPCR analysis of transgenic *Nanog* expression 24 h after Dox addition in male/female GOF-GFP ES cells. $\Delta\Delta C_t$ mean values \pm s.d.; $n = 4$ values obtained from two technical replicates from each of two biological replicates. Two-sided/unpaired t -test: $**P < 0.01$. Related to Fig. 1a. **b**, Representative brightfield/GFP images of male day-2 PGCLCs induced from GOF-GFP day-2 EpiLCs; plus Dox for *Nanog* expression. Scale bar, 200 μ m. Related to Fig. 1a. **d**, day 4. **c**, Representative FACS analysis of male day-4 PGCLCs (shown in Fig. 1a) induced from GOF-GFP day-2 EpiLCs; plus Dox for *Nanog* expression. SSC, side scatter. **d**, Representative brightfield/GFP images of female day-2 and day-4 PGCLCs induced from

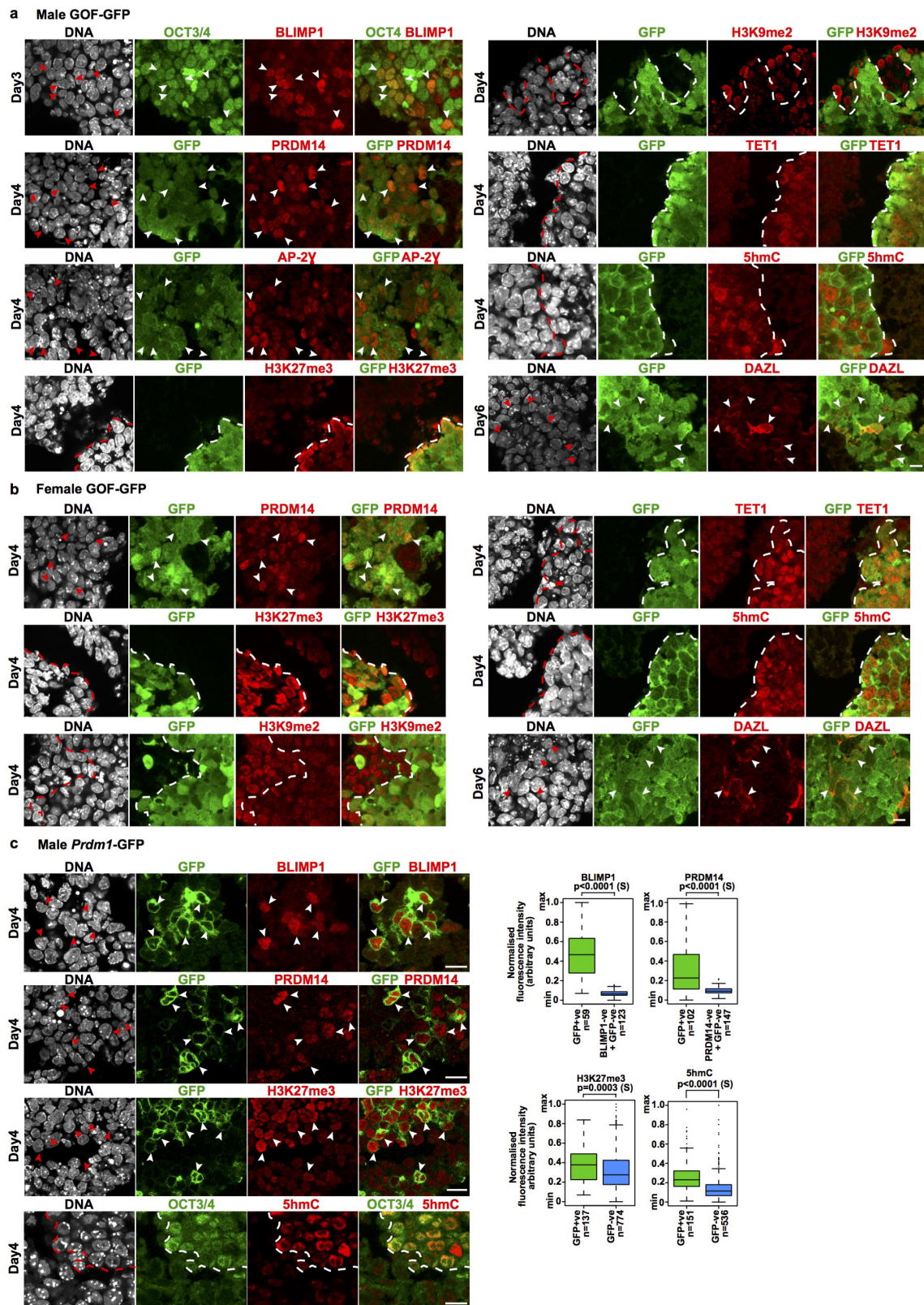
GOF-GFP day-2 EpiLCs; plus Dox for *Nanog* expression. Scale bar, 200 μ m. Related to Fig. 1a. **e**, FACS analysis for GFP with samples shown in **d**. Related to Fig. 1a. **f**, qPCR analysis of transgenic *Oct3/4* expression 24 h after Dox addition in male ES cells. $\Delta\Delta C_t$ mean values \pm s.d.; $n = 4$ values obtained from two technical replicates from each of two biological replicates. Two-sided/unpaired t -test: $**P < 0.01$. **g**, Expression of *Oct3/4* (unlike *Nanog*) does not result in the induction of GFP⁺ cells. PGCLC induction from female GOF-GFP EpiLCs; plus Dox for *Oct3/4* or *Nanog* expression. Representative brightfield/GFP images at day 4. Scale bar, 200 μ m. **h**, FACS analysis for GFP with samples shown in **g**.



Extended Data Figure 3 | See next page for caption.

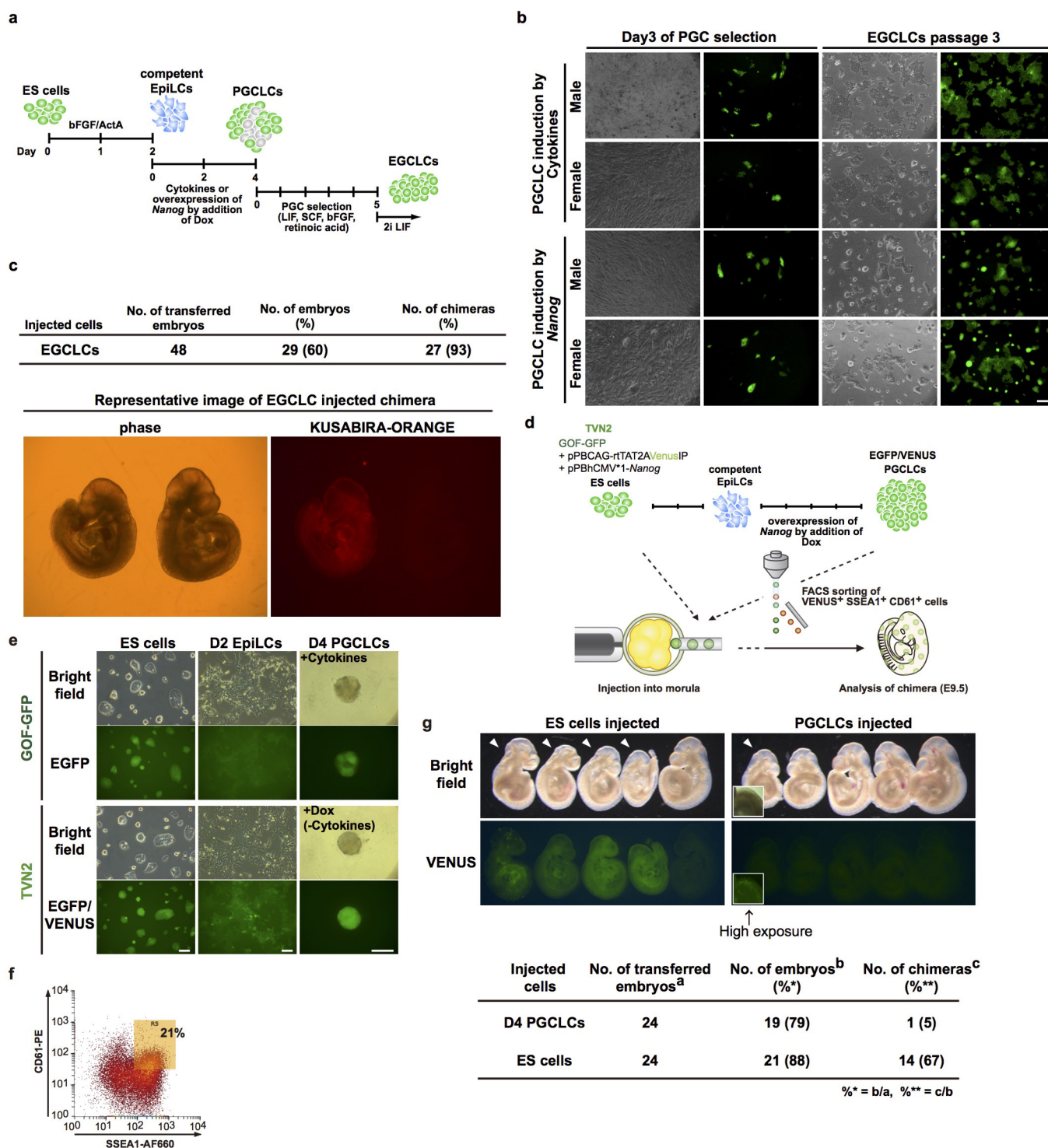
Extended Data Figure 3 | The transcriptomes of day-4 *Nanog*- and cytokine-induced PGCLCs are highly similar. **a**, 100–200 ng ml⁻¹ of Dox in EpiLCs results in NANOG expression levels similar to ES cells as shown by western blot analysis for NANOG and α -tubulin (α -TUB) with GOF-GFP ES cells and day-2 EpiLCs 24 h after PGCLC induction (EpiLC aggregations) with *Nanog* (+Dox). For gel source data, see Supplementary Fig. 1. **b**, PGCLC induction with 100 or 700 ng ml⁻¹ Dox (for *Nanog* expression) with noggin from GOF-GFP EpiLCs. Representative brightfield/GFP images at day 4. GFP⁺ cells are induced in both conditions. Scale bar, 200 μ m. **c**, Physiological (equivalent to ES cells) or higher levels of *Nanog* induce PGCLCs with comparable efficiency. FACS for GFP at day 4 of PGCLC induction with 100 or 700 ng ml⁻¹ Dox (for *Nanog* expression) with noggin from GOF-GFP or *Prdm1*-GFP EpiLCs. D, day; SSC, side scatter. **d**, Alternative representation of qPCR data for *Nanog*, *Prdm1* and *Tfap2c* shown in Fig. 1c. The induction of these genes in plus cytokine conditions appears less evident, when compared with plus Dox conditions. The data were log₂-scaled, which allows a better comparison. **e**, qPCR analysis of female GOF-GFP cells. GFP⁺ cells were FACS-sorted. Note the upregulation of PGC markers but not of the

ES-cell marker *Klf4*. $\Delta\Delta C_t$ mean values \pm s.d.; $n = 3$ biological replicates. Colour code is shown in **d**. Related to Fig. 1c. **f**, qPCR analysis of male *Prdm1*-GFP cells. GFP⁺ cells were FACS-sorted. Note the upregulation of PGC markers but not of the ES-cell marker *Klf4*. $\Delta\Delta C_t$ mean values \pm s.d.; $n = 3$ biological replicates. Colour code is shown in **d**. Related to Fig. 1c. **g**, The transcriptomes of *Nanog*- and cytokine-induced PGCLCs are highly similar. Scatter plot showing the correlation of microarray data of ES cells, FACS-sorted day-4 PGCLCs induced by cytokines or *Nanog* with noggin. R indicates the Pearson correlation coefficient. $n = 2$ biological replicates; related to Fig. 1d. **h**, *Nanog*- and cytokine-induced PGCLCs cluster together as shown in unsupervised hierarchical clustering of microarray data described in **g**. Related to Fig. 1d. **i**, Heat map showing the expression levels of selected genes from microarray data described in **g**. Related to Fig. 1d. **j**, *Nanog*-induced day-4 PGCLCs are closely related to cytokine-induced day-6 PGCLCs. PCA analysis with published microarray data sets⁵ (cross-platform comparison; see Methods for details). Note that the separation of ES cell samples is probably due to differences in genomic background and culture conditions.



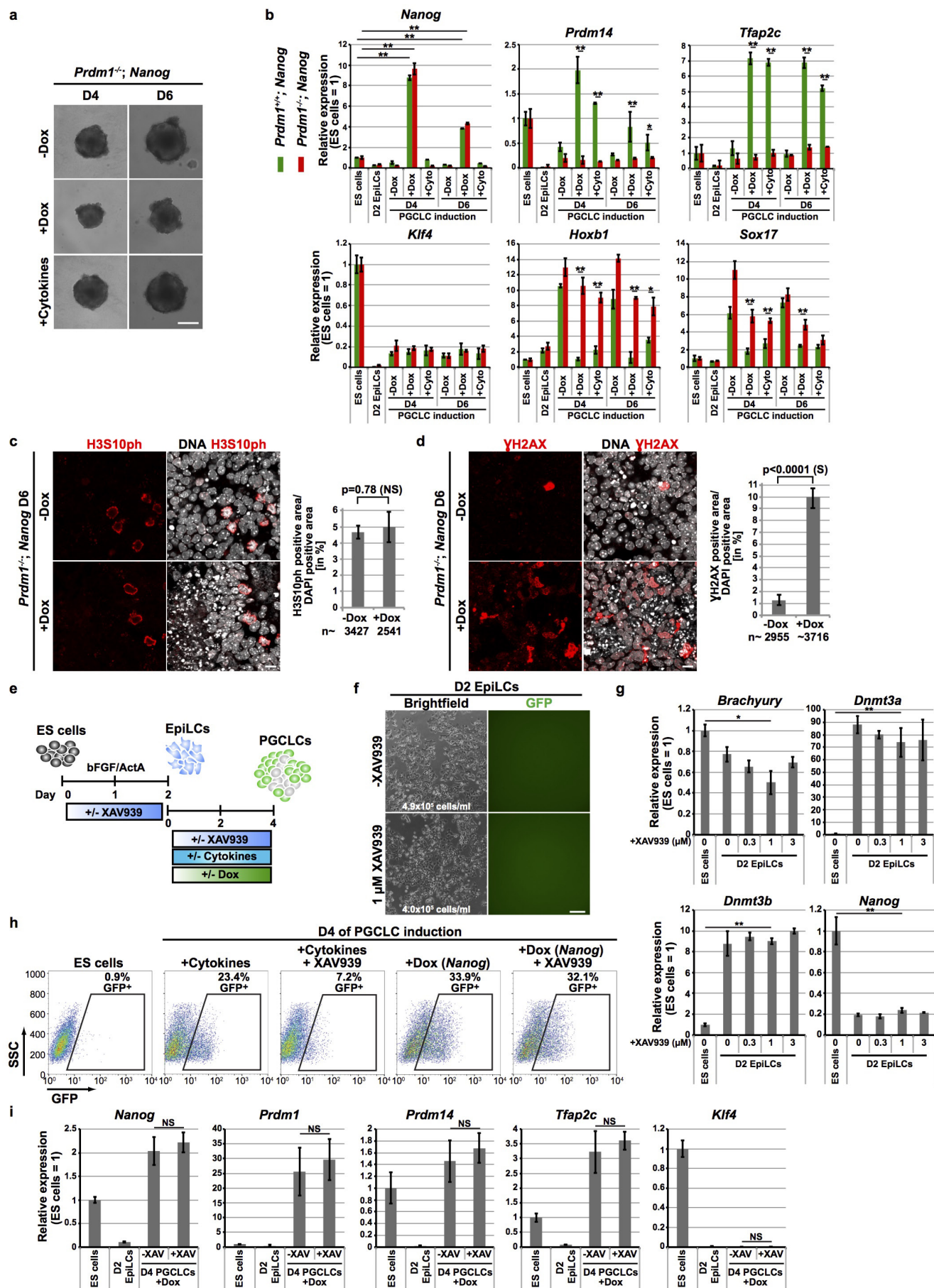
Extended Data Figure 4 | *Nanog*-induced PGCLCs show hallmarks of PGC development. a–c, Immunofluorescence analysis of PGC markers in GFP⁺ cells induced by *Nanog* from male (a) and female (b) GOF-GFP and male *Prdm1*-GFP (c) EpiLCs shows expression of BLIMP1, PRDM14, AP-2 γ and TET1, enrichment of H3K27me3 and 5hmC and a decrease of

H3K9me2 intensity; DAZL is detected in some cells on day 6. Arrowheads and dashed lines highlight single or cluster of GFP⁺ cells. $n = 2$ biological replicates. Scale bar, 10 μ m. Quantification in c was scale normalized. Two-sided/unpaired t -test: S, significant ($P \leq 0.01$); n = number of cells analysed. Related to Fig. 1e.



Extended Data Figure 5 | Functional analysis of *Nanog*-induced PGCLCs. **a**, Experimental design (for **b**, **c**) for the derivation of EGCLCs. PGCLCs were induced with cytokines or by *Nanog* (+Dox) from male or female GOF-GFP EpiLCs carrying a constitutively active Kusabira-Orange reporter. On day 4, aggregations were dissociated and cultured on mitomycin C-treated mouse embryonic fibroblast (MEF) feeder cells in PGC selection medium (LIF, SCF, bFGF, retinoic acid) for 5 days. After the selection, selected colonies were dissociated and transferred into ES-cell medium (2i/LIF). **b**, Experiment was performed as shown in **a**. Left panel shows representative images of proliferating GFP⁺ cells after 3 days of PGC selection. Right panel shows established EGCLCs after three passages in 2i/LIF. **c**, EGCLCs derived from day-4 PGCLCs by *Nanog* expression were injected into blastocysts, resulting in high contribution to chimaeras

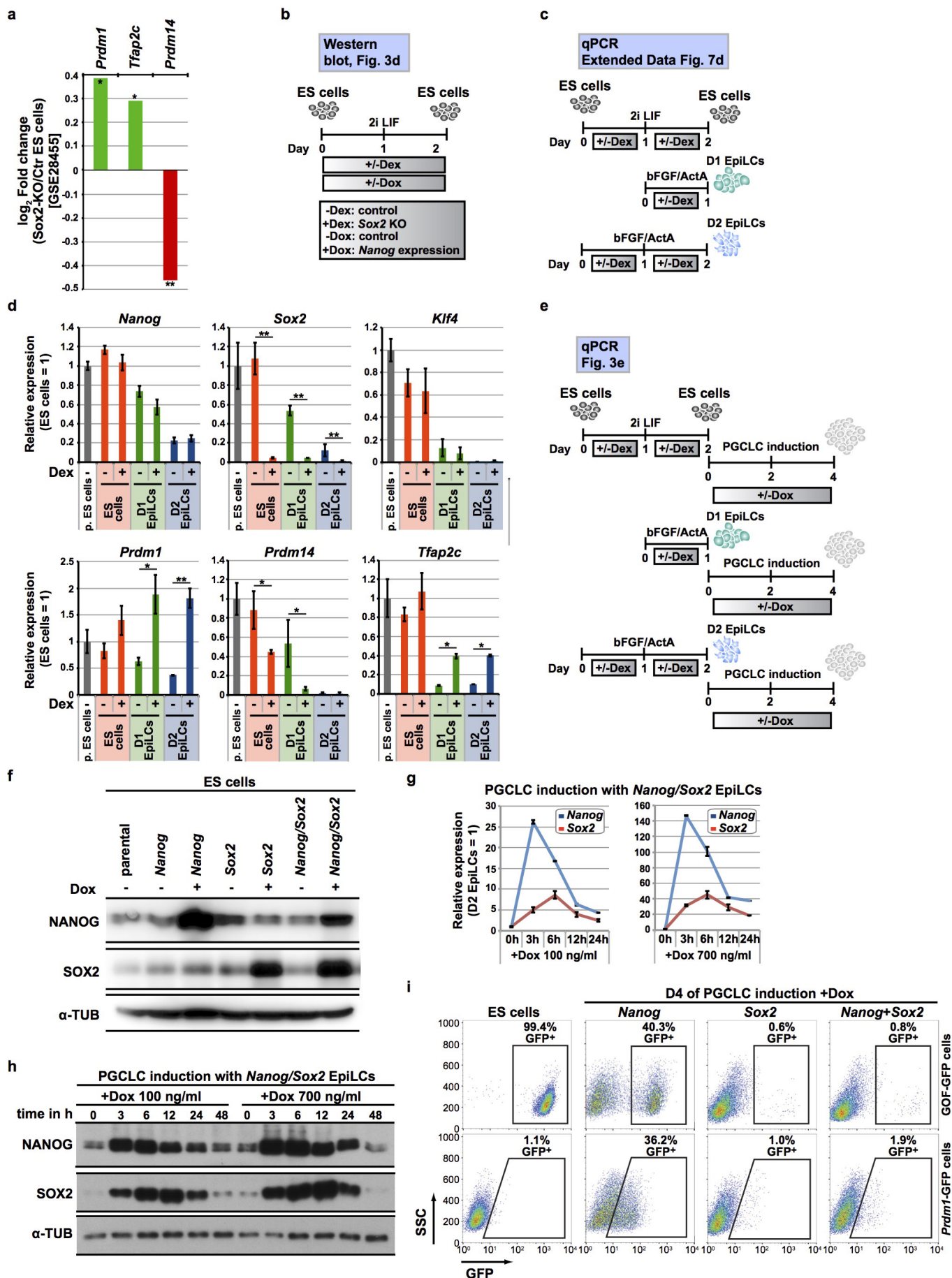
at E9.5 as shown by Kusabira-Orange expression. **d**, Experimental design (for **e**–**g**) for generating chimaeras. PGCLCs were induced from a GOF-GFP ES cell line expressing a fluorescent VENUS reporter constitutively and *Nanog* upon Dox addition (TVN2 cell line). On day 4, aggregations were dissociated and SSEA1⁺ and CD61⁺ cells were sorted by FACS, injected into morulae and analysed on E9.5. **e**, Representative brightfield, GFP/VENUS images of GOF-GFP or TVN2 cells during PGCLC induction by cytokines or *Nanog* (+Dox). Scale bars, 100 μm. **d**, day 4. **f**, FACS profile for SSEA1⁺ and CD61⁺ PGCLCs on day 4 induced as described and shown in **d**, **e**. **g**, ES cells but not PGCLCs contribute efficiently to chimaeras. ES cells or FACS-sorted *Nanog*-induced SSEA1⁺/CD61⁺ PGCLCs at day 4 were injected into morulae and representative brightfield/VENUS images from chimaeras at E9.5 are shown.



Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | *Prdm1*^{-/-} abrogates PGCLC induction by *Nanog*; *Nanog* and the WNT pathway act independently. **a**, PGCLC induction with cytokines or *Nanog* (+Dox) from *Prdm1*^{-/-} ES cells (*Prdm1*^{-/-}; *Nanog*). Representative brightfield images of day-4 and day-6 aggregations. Scale bar, 200 μ m. Related to Fig. 2a. **b**, Loss of *Prdm1* abrogates PGCLCs induced by NANOG as shown by qPCR analysis of mutant (*Prdm1*^{-/-}; *Nanog*) compared with control (*Prdm1*^{+/+}; *Nanog*) cells with *Nanog* (+Dox) or cytokines (+cyto). Unsorted samples were used for analysis. Note that the data shown in Fig. 2a was combined with additional qPCR data on cells at day 6 of PGCLC induction. $\Delta\Delta C_t$ mean values \pm s.d.; $n = 4$ values obtained from two technical replicates from each of two biological replicates. Two-sided/unpaired t -test: $**P < 0.01$; $*P < 0.05$. Related to Fig. 2a. **c**, *Nanog* does not affect cell proliferation rate of *Prdm1*^{-/-}; *Nanog* cells. Immunofluorescence staining for the mitotic marker H3S10ph in *Prdm1*^{-/-}; *Nanog* cells at day 6 of PGCLC induction; plus Dox for *Nanog* expression. Scale bar, 10 μ m. Two-sided/unpaired t -test; NS, not significant ($P > 0.01$); n = estimated number of cells (see Methods for details). Related to Fig. 2a. **d**, Induced expression of *Nanog* results in an increased number of cell death of *Prdm1*^{-/-}; *Nanog* cells. Immunofluorescence stainings of the DNA double-strand-break marker γ H2AX in *Prdm1*^{-/-}; *Nanog* cells at day 6 of PGCLC induction; plus Dox for *Nanog* expression. Scale bar, 10 μ m. Two-sided/unpaired t -test; s, significant ($P \leq 0.01$); n = estimated number of cells (see Methods for details). Related to Fig. 2a. **e**, Experimental design (for **f**–**i**) to test the

interdependence of *Nanog* and the WNT pathway for PGCLC induction. *Prdm1*-GFP ES cells were sequentially differentiated into PGCLCs plus or minus tankyrase inhibitor XAV939, which causes the degradation of β -catenin²⁵; plus Dox for *Nanog* expression. **f**, XAV939 does not affect the morphology and proliferation of day-2 EpiLCs. Representative brightfield/GFP images of day-2 EpiLCs induced from GOF-GFP ES cells with 1 μ M XAV939. Scale bar, 200 μ m. **g**, qPCR of day-2 EpiLCs treated with XAV939 as shown in **e**, **f**. The expression of *Nanog* and of the EpiLC markers *Dnmt3a* and *Dnmt3b* are not affected by XAV939. *Brachyury*, the downstream target of WNT, is most efficiently repressed with 1 μ M XAV939. $\Delta\Delta C_t$ mean values \pm s.d.; $n = 4$ values obtained from two technical replicates from each of two biological replicates. Two-sided/unpaired t -test: $**P < 0.01$; $*P < 0.05$. **h**, The efficiency of PGCLC induction by cytokines but not by *Nanog* (+Dox) is markedly reduced upon XAV939 addition. PGCLCs were induced from 1 μ M XAV939-treated day-2 EpiLCs. Representative FACS analysis for GFP with cells at day 4 of PGCLC induction. SSC, side scatter. **i**, XAV939 does not affect the induction of PGC marker expression in *Nanog*-induced PGCLCs. Gene expression analysis by qPCR with FACS-sorted *Nanog*-induced day-4 PGCLCs plus or minus 1 μ M XAV939. Mean $\Delta\Delta C_t$ values \pm s.d.; $n = 4$ values obtained from two technical replicates from each of two biological replicates. Two-sided/unpaired t -test: $**P < 0.01$; $*P < 0.05$; NS, not significant.



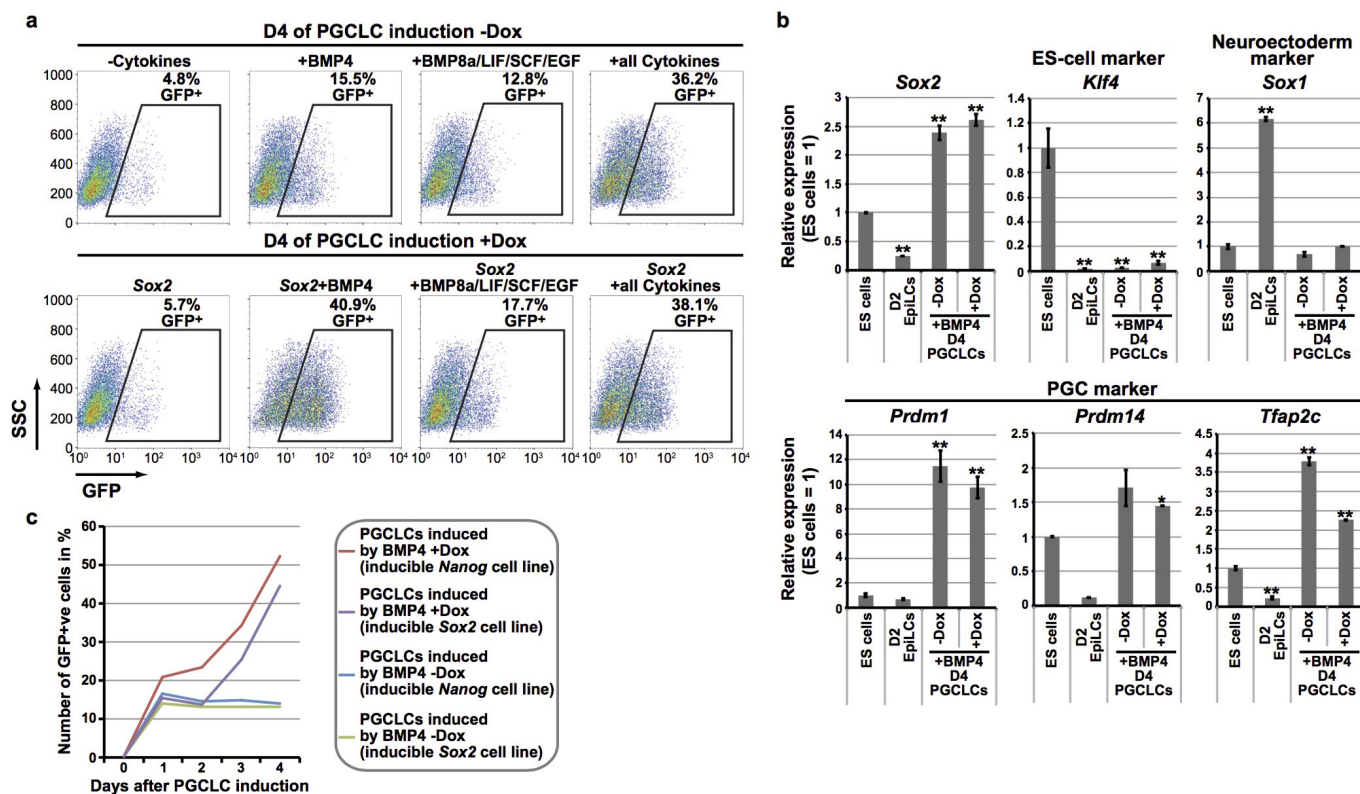
Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Sox2 inhibits PGCLC induction by Nanog.

a, *Prdm1* and *Tfap2c* are upregulated and *Prdm14* is downregulated in *Sox2*-knockout (KO) ES cells from published microarray data²⁶.

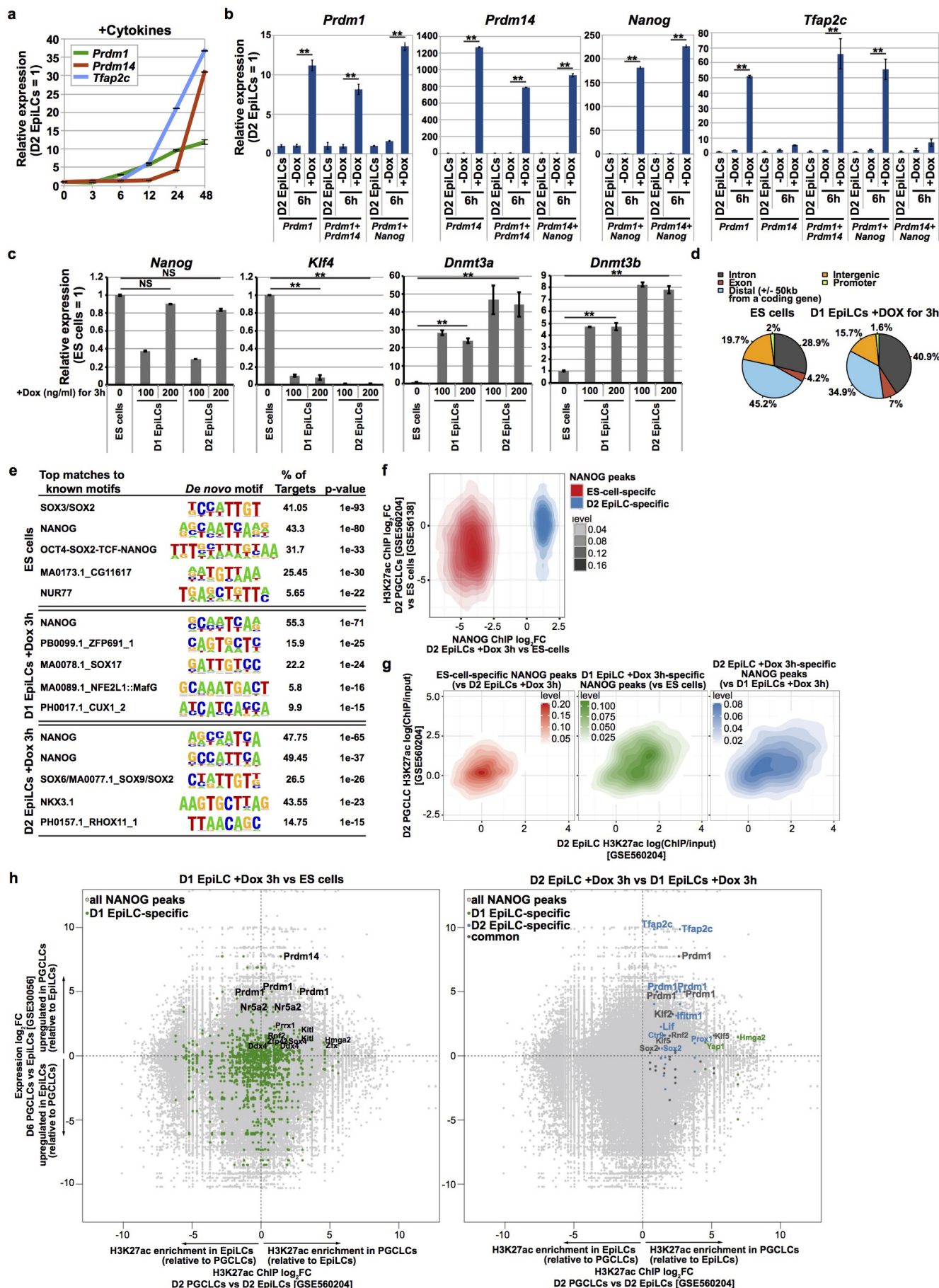
*****P* < 0.01; **P* < 0.05.** **b**, Experimental design for the western blot shown in Fig. 3d. Conditional *Sox2*-knockout ES cells carrying transgenes for Dox-inducible *Nanog* expression were treated with Dex to induce a *Sox2*-knockout and/or Dox for *Nanog* expression for 2 days. **c**, Experimental design for the qPCR analysis shown in **d**. *Sox2*-knockout ES cells: conditional *Sox2*-knockout ES cells carrying transgenes for Dox-inducible *Nanog* expression were treated plus or minus Dex for 2 days; *Sox2*-knockout day-1 EpiLCs: ES cells were cultured in 2i/LIF medium with Dex for 1 day and in bFGF/activin A (ActA) medium with Dex for one more day; *Sox2*-knockout day-2 EpiLCs: ES cells were transferred into bFGF/activin A medium containing Dex for 2 days. **D**, day. **d**, Loss of *Sox2* results in upregulation of *Prdm1* and *Tfap2c* and downregulation of *Prdm14* in ES cells, day-1 and day-2 EpiLCs; qPCR analysis following *Sox2*-knockout (+Dex). $\Delta\Delta C_t$ mean values \pm s.d.; *n* = 4 values obtained from two technical replicates from each of two biological replicates. Two-sided/unpaired *t*-test: *****P* < 0.01; **P* < 0.05.** Experimental design

is shown in **c**, p., parental. Related to Fig. 3e. **e**, Experimental design for the qPCR analysis shown in Fig. 3e. *Sox2*-knockout ES cells, day-1 or day-2 EpiLCs were generated as described in **c**, and subsequently induced into PGCLCs plus or minus *Nanog* (+/-Dox). **f**, Western blot for NANOG, SOX2 and α -tubulin (α -TUB) in GOF-GFP ES cells carrying Dox-inducible transgenes for *Nanog*, *Sox2* or *Nanog/Sox2* (+Dox for 24 h). Related to Fig. 3f. For gel source data, see Supplementary Fig. 1. **g**, Time-course qPCR analysis showing *Nanog* and *Sox2* expression kinetics during PGCLC induction. PGCLCs were induced from GOF-GFP EpiLCs; +100 or 700 ng ml⁻¹ Dox for *Nanog/Sox2* expression. $\Delta\Delta C_t$ mean values \pm s.d.; *n* = 4 values obtained from two technical replicates from each of two biological replicates. Related to Fig. 3f. **h**, Time-course western blot for NANOG, SOX2 and α -tubulin (α -TUB) showing NANOG and SOX2 protein kinetics during PGCLC induction. PGCLCs were induced from GOF-GFP EpiLCs; +100 or 700 ng ml⁻¹ Dox for *Nanog/Sox2* expression. For gel source data, see Supplementary Fig. 1. Related to Fig. 3f. **i**, FACS analysis for GFP at day 4 of PGCLC induction from GOF-GFP or *Prdm1*-GFP EpiLCs; plus Dox for *Nanog*, *Sox2* or *Nanog/Sox2* expression. Related to Fig. 3f.



Extended Data Figure 8 | Sox2 positively affects cell proliferation rate of cytokine-induced PGCLCs. **a**, Sox2 increases the number of GFP⁺ cells induced by BMP4 alone. Representative FACS analysis for GFP at day 4 of PGCLC induction from *Prdm1*-GFP EpiLCs; plus Dox for Sox2 expression. D, day; SSC, side scatter. **b**, Sox2 does not affect the upregulation of PGC markers in cytokine-induced PGCLCs. qPCR analysis of FACS-sorted GFP⁺ cells induced by BMP4 and/or Sox2 (+Dox). $\Delta\Delta C_t$ mean values \pm s.d.; $n = 4$ values obtained from two technical replicates from each of

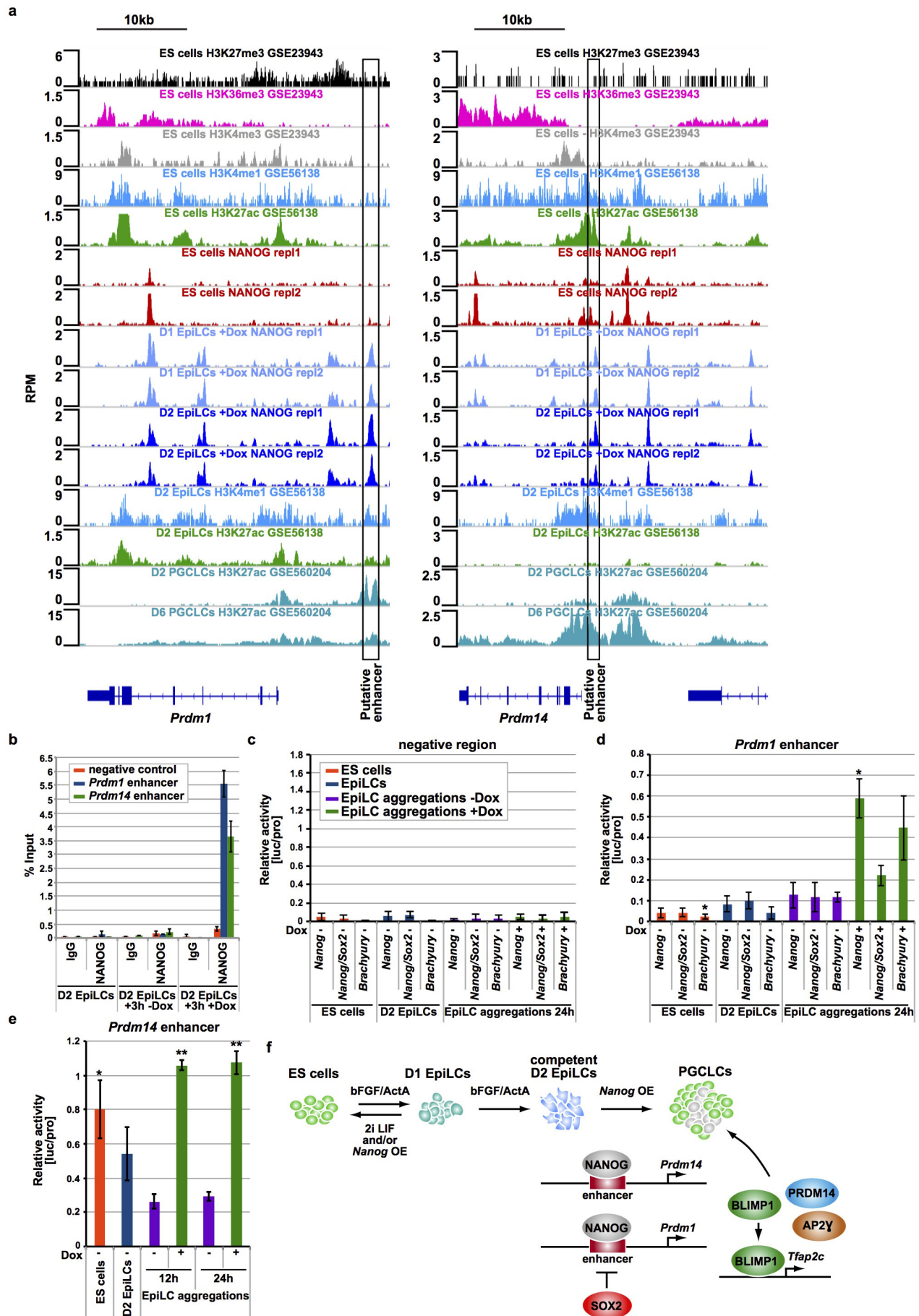
two biological replicates. ES cells were used as a reference for P values (two-sided/unpaired t -test): $**P < 0.01$; $*P < 0.05$. **c**, Time-course FACS analysis of GFP⁺ cells after PGCLC induction with BMP4 and plus or minus *Nanog* or *Sox2* (+/-Dox). The number of GFP⁺ cells at day 2 of PGCLC induction with or without Sox2 expression is comparable, but increased with *Nanog*. After day 2, PGCLCs induced by BMP4 with Sox2 or *Nanog* increase their proliferation rate.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | *Nanog* shows a cell-type-specific binding pattern and induces *Prdm1*, *Prdm14* and *Tfap2c*. **a**, Time-course qPCR for *Prdm1*, *Prdm14* and *Tfap2c* between 1–48 h after PGCLC induction with cytokines from GOF-GFP EpiLCs. $\Delta\Delta C_t$ mean values \pm s.d.; $n = 3$ technical replicates. Related to Fig. 4a. **b**, *Prdm1* alone can induce the expression of *Tfap2c*. GOF-GFP EpiLCs with combinations of Dox-inducible transgenes encoding *Prdm1*, *Prdm14* and/or *Nanog* plus or minus Dox for 6 h were analysed by qPCR. The expression of *Prdm1*, *Prdm14* and/or *Nanog* is upregulated in the corresponding EpiLCs upon Dox addition. $\Delta\Delta C_t$ mean values \pm s.d.; $n = 4$ values obtained from two technical replicates from each of two biological replicates. Two-sided/unpaired *t*-test: $**P < 0.01$; $*P < 0.05$. **D**, day. **c**, To acquire sufficient numbers of cells for ChIP-seq studies, GOF-GFP day-1 or day-2 EpiLCs ($\sim 1 \times 10^6$ cells per 6-cm plate) with Dox-inducible *Nanog* transgenes were aggregated in low-binding plates plus Dox to induce PGCLCs. qPCR analysis of day-1 and day-2 EpiLCs after 3 h with 100 or 200 ng ml⁻¹ Dox is shown. The addition of 200 ng ml⁻¹ of Dox results in *Nanog* expression levels comparable to ES cells after 3 h. $\Delta\Delta C_t$ mean values \pm s.d.; $n = 4$ values obtained from two technical replicates from each of two biological replicates. Two-sided/unpaired *t*-test: $**P < 0.01$; NS, not significant. **D**, day. **d**, NANOG ChIP-seq analysis shows genomic distribution of NANOG in GOF-GFP ES cells and day-1 EpiLCs plus *Nanog* (+Dox) for 3 h. ‘Distal’ refers to intergenic peaks, which are within ± 50 kb of an annotated coding

gene, while those further away are categorized as ‘intergenic’. Related to Fig. 4b. **e**, *De novo* motif analysis with NANOG ChIP-seq data. Shown are the top five matches of the *de novo* motifs to known motifs. The analysed cell types show enrichment for the NANOG and SOX motifs. ES cells show additional enrichment for pluripotency motifs, while EpiLCs show a different set of motif enrichment. **f**, Day-2 EpiLC-specific (D2 EpiLCs +Dox 3 h) NANOG-bound enhancers become more enriched for H3K27ac than ES-cell-specific NANOG-bound enhancers in cytokine-induced day-2 PGCLCs as compared to ES cells. Contour plots showing differential binding of NANOG in day-2 EpiLCs versus ES cells (*x*-axis) compared to the differential enrichment of H3K27ac in day-2 PGCLCs³⁰ versus ES cells⁹ (*y*-axis). **g**, NANOG binds enhancers that are enriched for H3K27ac in day-1/2 EpiLCs (D1/2 EpiLCs +Dox 3 h). A subset of enhancers, however, becomes more enriched for H3K27ac in cytokine-induced day-2 PGCLCs³⁰ as compared to day-2 EpiLCs³⁰. **h**, NANOG might contribute to the activation of enhancers associated with germline genes. Scatter plots show differential gene expression analysis between day-6 PGCLCs⁵ and day-2 EpiLCs⁵ (*y*-axis), and differential H3K27ac enrichment between day-2 PGCLCs³⁰ and day-2 EpiLCs³⁰ (*x*-axis) on NANOG-binding sites. The top 40% of NANOG peaks were associated with the nearest gene in a 200-kb window. Highlighted are candidate enhancers, which are associated with germline genes and become activated (H3K27ac-enriched) in PGCLCs.



Extended Data Figure 10 | See next page for caption.

Extended Data Figure 10 | *Nanog* induces PGC-like fate: a model.

a, ChIP-seq data tracks^{9,30,39} at the *Prdm1* and *Prdm14* loci for NANOG in ES cells, day-1 and day-2 EpiLCs (EpiLCs were collected after 3 h with 200 ng ml⁻¹ Dox for *Nanog* expression). Boxed are putative enhancer elements. The *Prdm1* enhancer is enriched for H3K4me1 in day-2 EpiLCs and gains H3K27ac in PGCLCs. The *Prdm14* enhancer shows enrichment for H3K4me1 in ES cells and EpiLCs and becomes enriched for H3K27ac in ES cells and PGCLCs but not in EpiLCs. Note that these enhancer marks follow the expression pattern of *Prdm1* or *Prdm14*, respectively. D, day; RPM, reads per million. Related to Fig. 4d. **b**, ChIP-qPCR validation of NANOG ChIP-seq data with GOF-GFP day-2 EpiLCs before and 3 h after PGCLC induction by *Nanog* expression (+Dox). NANOG is enriched at putative enhancer regions, which are close to *Prdm1* and *Prdm14*. Error bars indicate s.d.; $n = 4$ values obtained from two technical replicates from each of two biological replicates. **c**, ES cell lines with luciferase reporter plasmids with a genomic region, which does not show any enhancer marks and NANOG binding, and indicated Dox-inducible transgenes served as a negative control. Luciferase activity, measured in ES cells, day-2 EpiLCs and 24 h after PGCLC induction (EpiLC aggregations), was normalized to protein quantity (luc/pro). Mean values \pm s.d.; $n = 6$ values obtained from two technical replicates from each of two biological replicates. **d**, Biological

replicate experiment for the luciferase assay with the *Prdm1* enhancer as shown and described in Fig. 4e. Luciferase activity, measured in ES cells, day-2 EpiLCs and 24 h after PGCLC induction, was normalized to protein quantity (luc/pro). Mean values \pm s.d.; $n = 3$ technical replicates. Colour code is shown in **c**. Reference for P values (two-sided/unpaired t -test): EpiLC aggregations minus Dox; $**P < 0.01$; $*P < 0.05$. **e**, Biological replicate experiment for the luciferase assay with the *Prdm14* enhancer as shown and described in Fig. 4g. Luciferase activity, measured in ES cells, day-2 EpiLCs and 12/24 h after PGCLC induction, was normalized to protein quantity (luc/pro). Mean values \pm s.d.; $n = 3$ technical replicates. Colour code is shown in **c**. Reference for P values (two-sided/unpaired t -test): EpiLC aggregations minus Dox 24 h; $**P < 0.01$; $*P < 0.05$. **f**, Model showing the role of NANOG during PGCLC induction *in vitro*. Day-1 EpiLCs are not competent to become PGCLCs, but retain the capability to revert to an ES-like state via 2i/LIF and/or *Nanog* overexpression. Day-2 EpiLCs differentiate into PGCLCs upon *Nanog* expression. NANOG binds to putative enhancer elements of *Prdm1* and *Prdm14* to activate their transcription, which is sufficient to induce the PGCLC fate. This effect can be antagonized by SOX2, which co-binds the *Prdm1* enhancer.

A phosphoinositide conversion mechanism for exit from endosomes

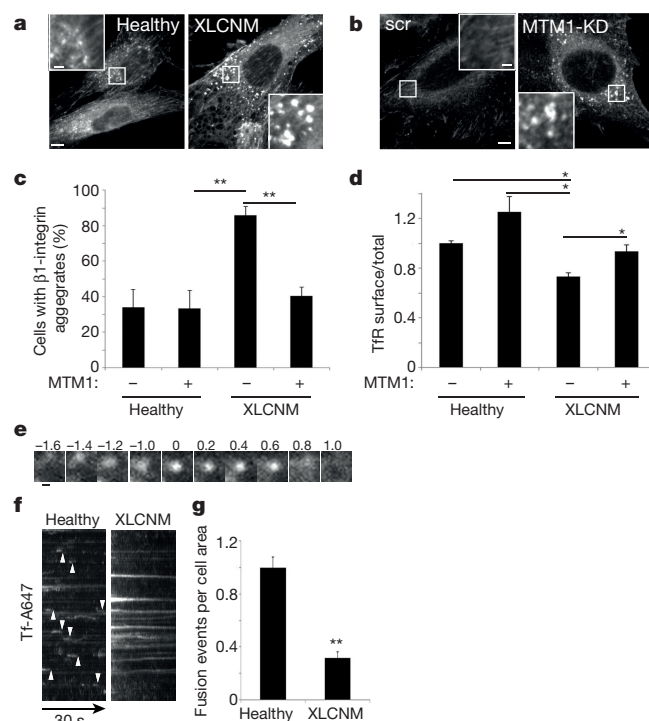
Katharina Ketel¹, Michael Krauss¹, Anne-Sophie Nicot², Dmytro Puchkov¹, Marnix Wieffer³, Rainer Müller⁴, Devaraj Subramanian⁴, Carsten Schultz⁴, Jocelyn Laporte² & Volker Haucke^{1,3,5}

Phosphoinositides are a minor class of short-lived membrane phospholipids that serve crucial functions in cell physiology ranging from cell signalling and motility to their role as signposts of compartmental membrane identity^{1,2}. Phosphoinositide 4-phosphates such as phosphatidylinositol 4-phosphate (PI(4)P) and phosphatidylinositol 4,5-bisphosphate (PI(4,5)P₂) are concentrated at the plasma membrane, on secretory organelles³, and on lysosomes⁴, whereas phosphoinositide 3-phosphates, most notably phosphatidylinositol 3-phosphate (PI(3)P)⁵, are a hallmark of the endosomal system^{1,2}. Directional membrane traffic between endosomal and secretory compartments, although inherently complex, therefore requires regulated phosphoinositide conversion. The molecular mechanism underlying this conversion of phosphoinositide identity during cargo exit from endosomes by exocytosis is unknown. Here we report that surface delivery of endosomal cargo requires hydrolysis of PI(3)P by the phosphatidylinositol 3-phosphatase MTM1, an enzyme whose loss of function leads to X-linked centronuclear myopathy (also called myotubular myopathy) in humans⁶. Removal of endosomal PI(3)P by MTM1 is accompanied by phosphatidylinositol 4-kinase-2 α (PI4K2 α)-dependent generation of PI(4)P and recruitment of the exocyst tethering complex to enable membrane fusion. Our data establish a mechanism for phosphoinositide conversion from PI(3)P to PI(4)P at endosomes en route to the plasma membrane and suggest that defective phosphoinositide conversion at endosomes underlies X-linked centronuclear myopathy caused by mutation of MTM1 in humans.

X-linked centronuclear myopathy (XLCNM) is a severe congenital myopathy caused by mutations in MTM1 (ref. 6). Patients with XLCNM suffer from defective muscle fibres due to disorganized membrane compartments and T-tubules⁷ and are affected by severe hypotonia at birth, while the mildest surviving cases display muscle and non-muscle phenotypes⁸. MTM1 knockout mice are born below Mendelian ratios, and show progressive myopathy and growth impairment leading to early death⁹.

Depletion of MTM1 in fruitflies causes endosomal accumulation of β 1-integrin in developing muscles¹⁰, possibly because of defective phosphoinositide conversion at endosomes. Intracellular β 1-integrin accumulations were indeed seen in endosomes of the MTM1-deficient XLCNM-patient fibroblast cell line H31 (Fig. 1a) and in MTM1-depleted HeLa cells (Fig. 1b and Extended Data Fig. 1a). Re-expression of active MTM1 rescued β 1-integrin accumulation in H31 cells (Fig. 1c). Accumulations of β 1-integrin in MTM1-depleted cells were immunopositive for recycling endosomal transferrin receptor (TfR) and late endosomal Rab7, but displayed limited overlap with early endosomal Rab5 and were negative for the trans-Golgi network marker Rab8 (Extended Data Fig. 1a). These data suggest that MTM1 loss-of-function causes β 1-integrin accumulation in late endosomes, possibly to facilitate their degradation. We hypothesized that defective

cargo exit from endosomes may underlie the endosomal sequestration of β 1-integrin in XLCNM-patient cells. Therefore, we analysed the distribution of TfR, a receptor that, following internalization, is recycled by exocytosis from endosomes to the plasma membrane¹¹. TfR surface levels, but not the total TfR levels, were decreased in H31 cells. This was rescued by re-expression of MTM1 (Fig. 1d and



¹Leibniz-Institut für Molekulare Pharmakologie, 13125 Berlin, Germany. ²Department of Translational Medicine and Neurogenetics, Institut de Genetique et de Biologie Moleculaire et Cellulaire (IGBMC), INSERM U964, CNRS UMR7104, Strasbourg University, 67404 Illkirch, France. ³Freie Universität Berlin, Faculty of Biology, Chemistry and Pharmacy, 14195 Berlin, Germany. ⁴European Molecular Biology Laboratory (EMBL), Cell Biology and Biophysics Unit, 69117 Heidelberg, Germany. ⁵NeuroCure Cluster of Excellence, Charité Universitätsmedizin Berlin, 10117 Berlin, Germany.

Extended Data Fig. 1b). Thus, MTM1 is required for efficient surface delivery of TfR, presumably by regulating exocytosis from endosomes. We directly monitored Tf exocytosis from endosomes by total internal reflection fluorescence (TIRF) microscopy. In fibroblasts from healthy controls, Tf-containing endosomal structures approached the plasma membrane before exocytic release of their Tf content. Exocytic fusion events were detected much less frequently in H31 cells (Fig. 1e–g), where Tf-loaded endosomes were largely immobile, a phenotype confirmed in cells from another patient with XLCNM (Extended Data Fig. 1c).

To dissect MTM1 function, we knocked down its expression in HeLa cells. MTM1 loss led to a partial depletion of TfR from the plasma membrane, resulting in reduced Tf internalization upon MTM1 knockdown (MTM1-KD) that was proportional to the diminished TfR surface levels (Fig. 2a and Extended Data Fig. 2a, b). Internalization kinetics and TfR expression were unaltered (Extended Data Fig. 2c–f). Thus, MTM1 is required for efficient surface delivery of cargo by mediating exocytosis from endosomes. Consistent with this hypothesis, exocytic fusion events were detected much less frequently in MTM1-depleted HeLa cells, where stalled endosomes that failed to fuse accumulated underneath the plasma membrane, a phenotype rescued by re-expression of active but not phosphatase-inactive (C375S)¹² MTM1 (Fig. 2b and Extended Data Fig. 2g–j). Stalled TfR-containing endosomes in MTM1-KD cells were accessible to internalized Tf ligand and distinct from surface TfR pools (Fig. 2c and Extended Data Fig. 2i–j). MTM1 depletion did not significantly affect degradative sorting of internalized epidermal growth factor (EGF), EGF receptor levels, secretion, or retrograde traffic of endocytosed cholera toxin B; nor did we detect overt effects on the number of LC3B-containing autophagosomes (Extended Data Fig. 3a–h). These data suggest that MTM1 depletion does not lead to a general impairment of endosome function.

In the absence of MTM1, endosomal TfR was redistributed to the cell periphery or remained scattered throughout the cytoplasm instead of its usual perinuclear concentration (Fig. 2c, d and Extended Data Fig. 2j). A similar, although less severe, phenotype was seen in cells depleted of the related phosphoinositide 3-phosphatase MTMR1 (ref. 6), while knockdown of other MTMs was without effect (Extended Data Fig. 3i–k). To investigate the effect of XLCNM-patient mutations on exocytosis, we re-expressed active MTM1, phosphatase-inactive MTM1 (C375S), or MTM1 constructs with XLCNM mutations in MTM1-depleted HeLa cells. Phosphatase-inactive or disease-associated MTM1 mutants were unable to restore normal TfR-endosome distribution (Fig. 2e). These data identify MTM1 as a key factor for endosomal exocytosis, while defective exocytosis results from MTM1 dysfunction in cells from patients with XLCNM.

We characterized the nature of sub-plasma-membrane endosomes accumulating in MTM1-depleted cells. MTM1-KD cells incubated with horseradish-peroxidase-coupled Tf displayed an accumulation of tubular and horseshoe-shaped organelles typical of early and recycling endosomes¹¹ (Extended Data Fig. 2k–m). TfR-positive endosomes often formed elongated membrane tubules (Fig. 2c and Extended Data Fig. 4a) and contained early and recycling endosomal markers such as early endosomal antigen 1 (EEA1), Rab4, Rab5, Rab11, and Rab14 (ref. 11) (Fig. 3a and Extended Data Fig. 4b–e), and endosomal sorting nexins (SNXs) SNX1, SNX3, SNX4, SNX8, SNX17, and SNX27 (ref. 13) (Fig. 3a and Extended Data Fig. 5a, b). Importantly, loss of MTM1 did not interfere with the segregation of endosomal compartments marked by distinct Rab proteins (Extended Data Fig. 4f, g). TfR-endosomes were negative for the Golgi marker GM130, endosomal APPL1, late endosomal/lysosomal LAMP1, retromer, the endocytic adaptor AP-2, Rab35, a marker for fast recycling, or SNX15 (Extended Data Figs 4b–d and 5a, b). In agreement with MTM1 acting on endosomes¹², MTM1 co-localized with early and recycling endosomal markers, but much less with late endosomal or trans-Golgi network proteins (Extended Data Fig. 5c, d).

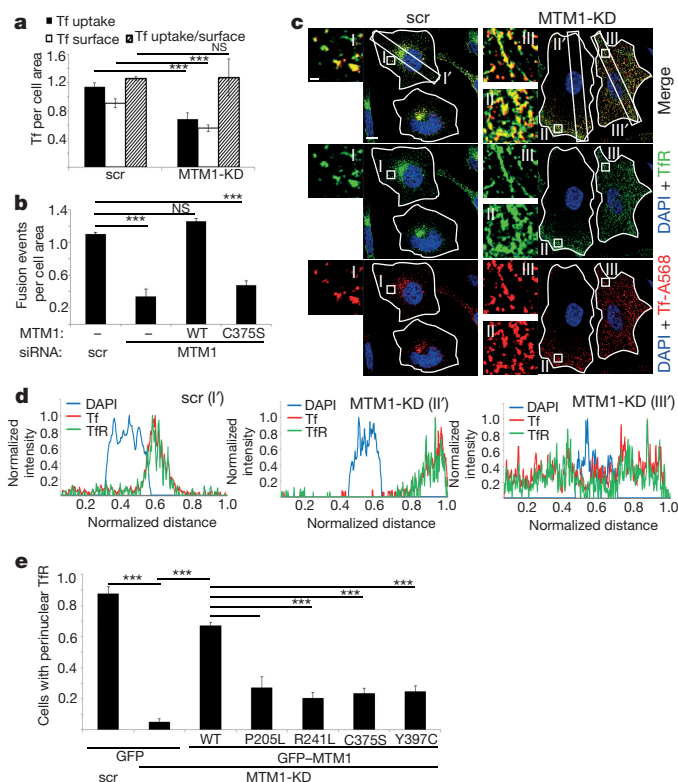


Figure 2 | Tf exocytosis from endosomes requires MTM1-mediated PI(3)P hydrolysis. **a**, Tf uptake, surface Tf levels, and ratio of uptake/surface-bound Tf in HeLa cells depleted of MTM1 (MTM1-KD) or scrambled siRNA-treated controls (scr) ($n = 3$). **b**, Impaired Tf exocytosis in MTM1-KD versus scr controls. Re-expression of wild-type (WT) eGFP-MTM1, but not catalytically inactive MTM1 C375S or eGFP, rescues defective Tf exocytosis ($n = 3$). **c**, Confocal images revealing co-localization of internalized Tf and TfR in MTM1-depleted cells. Scale bar, 10 μm; magnified insets, 1 μm. **d**, Averaged line scans over the marked area illustrate perinuclear accumulation of TfR in controls (**c**, **d**: I') and accumulation of TfR at the cell periphery (**c**, **d**: II') or cytosolic dispersion (**c**, **d**: III') in MTM1-depleted cells. **e**, Re-expression of eGFP-MTM1 WT but not C375S or XLCNM-patient mutations (P205L, R241L, Y397C) restores perinuclear TfR localization in MTM1-depleted cells. Normalized fraction of cells with perinuclear TfR was quantified ($n = 3$). n , number of independent experiments with 15–30 images (**a**, **e**) or 5 videos (**b**) analysed per condition per experiment. Mean \pm s.e.m.; NS, non-significant, *** $P < 0.01$, **** $P < 0.001$, unpaired two-tailed t -test. **c**, Representative images from one of three independent experiments are shown.

As MTM1 is a PI(3)P 3-phosphatase that preferentially associates with PI(3)P (Extended Data Fig. 5e), we investigated whether the sub-plasma-membrane accumulation of endosomes was caused by defects in phosphoinositide homeostasis. Titration of enhanced green fluorescent protein (eGFP)-2xPYVE of Hrs allowed the semi-quantitative determination of PI(3)P levels in control and MTM1-depleted cells (Extended Data Fig. 6a). Specificity was confirmed by applying the phosphoinositide 3-kinase inhibitor wortmannin and MTM1 overexpression (Extended Data Fig. 6b, d). Loss of MTM1 resulted in a significant elevation of PI(3)P levels, whereas those of PI(4)P, PI(3,4)P₂, and PI(4,5)P₂ remained unaltered (Fig. 3b and Extended Data Fig. 6c), consistent with ref. 12. Endosomal PI(3)P accumulation in MTM1-depleted cells was also detected by the fluorescently labelled PX domain of p40 (Phox)¹⁴ (Extended Data Fig. 6e, f) and was further enhanced by co-depletion of MTM1 and endosomal PI(3)P effectors SNX4 or SNX17 (Extended Data Fig. 6g, h), suggesting that endosomal PI(3)P is partly masked by bound SNX4 and SNX17 and possibly other PI(3)P effectors. These data indicate that MTM1-mediated hydrolysis of PI(3)P on endosomes triggers loss of PI(3)P-associated proteins (that is, SNXs) as a prerequisite for exocytosis.

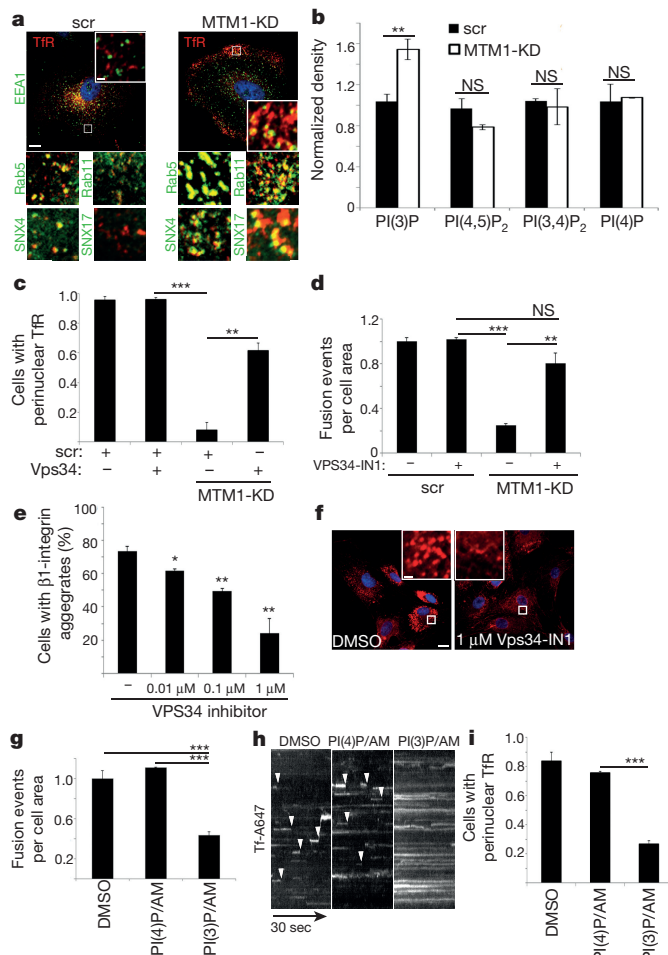


Figure 3 | Endosomal accumulation of PI(3)P and PI(3)P effector proteins inhibits exocytosis from endosomes. **a**, Distribution of endosomal markers in HeLa cells treated with scrambled (scr) siRNA or depleted of MTM1 (MTM1-KD). Confocal images stained for endogenous EEA1 or eGFP-Rab (Rab4, Rab5, Rab11, Rab14) or -SNX (SNX4, SNX8, SNX17, SNX27) proteins. **b**, Selective increase of PI(3)P in MTM1-depleted HeLa cells. Relative levels of PI(3)P, PI(4)P, PI(3,4)P₂, and PI(4,5)P₂ normalized to cell area are shown ($n = 3$). **c**, Co-depletion of MTM1 and Vps34 restores perinuclear TfR localization. Normalized fraction of cells with perinuclear TfR in siRNA-treated HeLa cells ($n = 3$). **d**, Inhibition of Vps34 (1 h, 37 °C) rescues defective Tf exocytosis in MTM1-depleted cells ($n = 3$). **e**, **f**, Inhibition of Vps34 rescues β 1-integrin accumulation in XLCNM patient cells. **e**, Percentage of cells with aggregated β 1-integrin ($n = 3$); **f**, β 1-integrin distribution in XLCNM patient cells. **g–i**, Exogenous oversupply of PI(3)P inhibits Tf exocytosis (**g**, Tf exocytic events per unit area ($n = 3$); **h**, kymographs of Tf fluorescence signal over 28 μ m, arrowheads indicate time of fusion) and causes TfR mislocalization (**i**). Normalized fraction of cells with perinuclear TfR is given ($n = 4$). **n**, number of independent experiments with 15–30 images (**b**, **c**, **e**, **i**) or 5 videos (**d**, **g**) analysed per condition per experiment. Mean \pm s.e.m.; NS, non-significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, unpaired two-tailed *t*-test. **a**, **f**, **h**, Representative images from one of three independent experiments are shown. **a**, **f**, Scale bar, 10 μ m; magnified insets, 1 μ m.

To challenge this hypothesis, we investigated whether defects induced by MTM1 depletion can be counteracted by manipulating PI(3)P synthesis⁵. Co-depletion of the phosphoinositide 3-kinase Vps34 together with MTM1 largely rescued defective TfR distribution seen in MTM1-KD cells (Fig. 3c and Extended Data Fig. 6i). Moreover, acute perturbation of PI(3)P synthesis by the selective Vps34-inhibitor Vps34-IN1 (ref. 15) restored defective Tf exocytosis and endosomal β 1-integrin accumulation in MTM1-KD or XLCNM-patient cells (Fig. 3d–f and Extended Data Fig. 6j). These data strengthen the close

correlation between impaired PI(3)P hydrolysis due to MTM1 loss-of-function and the strong disease pattern in patients with XLCNM, and open the possibility that defects in patients with XLCNM could be cured by pharmacological inhibition of Vps34 activity.

TfR dispersion in absence of MTM1 but not defective Tf exocytosis was also abrogated by knockdown of the PI(3)P-associated endosomal kinesin KIF16B (Extended Data Fig. 7a, b), indicating that defective Tf exocytosis is not a secondary consequence of endosome repositioning in the absence of MTM1.

In vitro MTM1 in addition to PI(3)P can also act on PI(3,5)P₂ (ref. 16). By depleting the PI(3)P-selective 5-kinase PIKfyve, we simultaneously interfered with endosomal PI(3,5)P₂ synthesis and PI(3)P consumption as a substrate¹. PIKfyve-KD did not restore perinuclear TfR-endosome localization or Tf exocytosis in MTM1-depleted cells. Instead, it mimicked the peripheral dispersion of TfR-endosomes seen in MTM1-KD cells, while co-depletion of both enzymes further aggravated endosomal defects evidenced by giant early endosome accumulation (Extended Data Fig. 7c–g). To directly probe whether oversupply of PI(3)P is sufficient to impair endosomal exocytosis, we applied cell-permeable PI-derivatives. Addition of PI(3)P¹⁷, but not PI(4)P, impaired Tf exocytosis and led to TfR-endosome dispersion (Fig. 3g–i). We conclude that MTM1-mediated hydrolysis of endosomal PI(3)P is a prerequisite for surface delivery of cargo by exocytosis from endosomes.

To address whether following PI(3)P hydrolysis endosomes have to acquire another phosphoinositide identity before exocytosis, we monitored their phosphoinositide content by dual-colour live TIRF microscopy. PI(3)P-positive endosomes inevitably failed to fuse. In contrast, endosomes undergoing productive exocytosis frequently acquired PI(4)P, which was lost concomitantly with Tf during the fusion process (Fig. 4a and Extended Data Fig. 8g). PI(3,4)P₂, PI(4,5)P₂, or PI(3,4,5)P₃ displayed no association with fusing endosomes. Thus, endosomal exocytosis of Tf appears to be mediated by PI(4)P on endosomes.

A candidate for endosomal PI(4)P synthesis is PI4K2 α , an enzyme localized to the trans-Golgi network¹⁸, endosomes^{1,19}, and secretory vesicles³ where it regulates protein sorting¹. MTM1 and PI4K2 α co-localized on endosomes and associated with each other (Fig. 4b, d and Extended Data Fig. 9a). In addition to myotubularin, quantitative mass spectrometry, together with affinity chromatography and co-immunoprecipitation, identified exocyst, an octameric PI(4)P-regulated complex²⁰ required for exocytosis, endosomal recycling²¹, and polarized sorting, as a binding partner of PI4K2 α (Fig. 4d and Extended Data Fig. 9a, b), and both proteins co-localized on endosomes (Fig. 4c). Furthermore, MTM1 associated with the Sec6 subunit of exocyst in yeast two-hybrid screens (not shown) and affinity chromatography (Extended Data Fig. 9c). Thus, MTM1, PI4K2 α , and exocyst may constitute a phosphoinositide conversion module for endosomal exocytosis.

Consistent with this hypothesis, depletion of PI4K2 α or exocyst caused defects in endosomal exocytosis and a partial dispersion of TfR, akin to MTM1-KD or MTM1/PI4K2 α double loss (Fig. 4e and Extended Data Fig. 9d–i). Exocyst recruitment to membranes is controlled by coincidence detection of PI(4)P and small GTPases²⁰, for example Sec4 in yeast and Rab11 in mammals²². Consistently, depletion of either PI4K2 α or Rab11a impaired both Tf exocytosis and exocyst recruitment to membranes (Fig. 4e–g and Extended Data Figs 8b–f and 10a–c). A PI(4)P requirement for exocyst recruitment was confirmed by overexpression of kinase-inactive PI4K2 α , and application of membrane-permeant PI(4)P (Extended Data Fig. 10d, e). Endosomes undergoing productive exocytosis contained Rab11, while those positive for Rab5 inevitably failed to fuse (Extended Data Fig. 8a). Taken together, PI(4)P produced by PI4K2 α on Rab11-endosomes is required for recruitment of exocyst to enable endosomal exocytosis.

Finally, we analysed the effects of PI4K2 α manipulation on membrane association of MTM1 and on PI(4)P and PI(3)P levels. Loss of PI4K2 α impaired membrane recruitment of MTM1 (Fig. 4h), while overexpression of active or inactive PI4K2 α facilitated MTM1

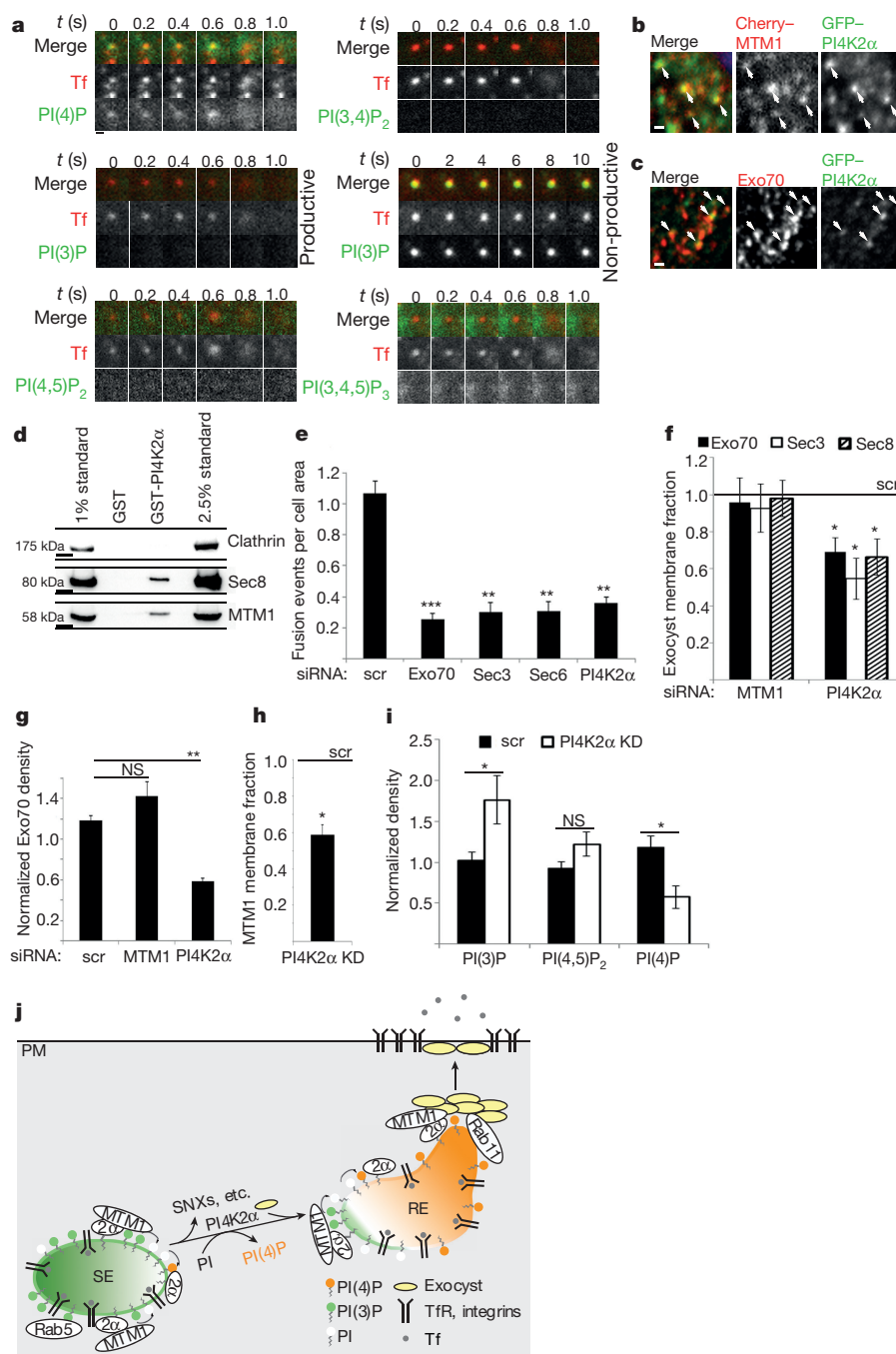


Figure 4 | PI(3)P to PI(4)P conversion is required for exocyst-dependent endosomal exocytosis. **a**, Phosphoinositide content of Tf exocytic vesicles monitored by dual-colour TIRF microscopy in HeLa cells expressing the corresponding phosphoinositide binding domains. Time course of representative Tf-containing endosomes is shown; scale bar, 400 nm. **b**, Co-localization of mCherry-MTM1 and eGFP-PI4K2 α on HeLa cell endosomes (arrows). **c**, Co-localization of eGFP-PI4K2 α and Exo70 on HeLa cell endosomes (arrows). **d**, GST-PI4K2 α associates with MTM1 and exocyst (Sec8). For blot source data, see Supplementary Fig. 1. **e**, Inhibition of Tf exocytosis in cells depleted of PI4K2 α - or exocyst subunits Exo70, Sec3, or Sec6 compared with scrambled (scr) siRNA-treated controls (Tf exocytic events per unit area; $n = 3$). **f–h**, Impaired membrane recruitment of exocyst and MTM1 in cells depleted of PI4K2 α . **f**, Relative membrane-associated fraction of exocyst subunits Exo70, Sec8, or Sec3 in scr control, MTM1-depleted, or PI4K2 α -depleted cells ($n = 5$).

g, Reduced endosomal exocyst association in PI4K2 α -depleted but not in MTM1-depleted cells revealed by confocal imaging ($n = 3$). **h**, Relative membrane-associated fraction of MTM1 in scr control or PI4K2 α -depleted cells ($n = 5$). **i**, Relative levels of PI(3)P, PI(4)P, and PI(4,5)P₂ in PI4K2 α -depleted HeLa cells ($n = 7$ for PI(3)P, $n = 3$ for all others). **j**, Model of phosphoinositide conversion during endosomal exocytosis. SE, sorting endosome; RE, recycling endosome; 2 α , PI4K2 α . n , number of independent experiments with one lysate (**f**, **h**), 15–30 images (**g**, **i**), or five videos (**e**) analysed per condition per experiment. Mean \pm s.e.m.; NS, non-significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, unpaired two-tailed t -test for immunocytochemistry (**g**, **i**) and live-TIRF (**e**), and one-sample t -test for western blot (**f**, **h**) quantifications. **a–c**, Representative images from one of four (**b**) or three (**a**, **c**) independent experiments are shown. **b**, **c**, Scale bar, 1 μ m.

association with endosomes, resulting in PI(3)P depletion (Extended Data Fig. 10f, g), consistent with the physical interaction of both proteins (Fig. 4d and Extended Data Fig. 9a). Furthermore, PI(3)P and

PI(4)P were enriched in distinct endosomal subdomains, suggesting that conversion of PI(3)P to PI(4)P may be initiated at Rab5-positive endosomes (Extended Data Fig. 10h). Importantly, PI4K2 α -KD cells

not only contained reduced levels of PI(4)P but also accumulated PI(3)P on endosomes containing PI(3)P binding effectors (Fig. 4i and Extended Data Fig. 10i–k) akin to MTM1 loss. Conversely, MTM1 was dispensable for the association of PI4K2 α with TfR-endosomes (Extended Data Fig. 10l), which is mediated by the endosomal SNARE cellubrevin/VAMP3 (ref. 23).

Our results reveal a mechanism for phosphoinositide conversion at endosomes from PI(3)P to PI(4)P to enable exit from the endosomal system. Phosphoinositide conversion is probably initiated at Rab5- and PI(3)P-containing early endosomes, where MTM1-mediated hydrolysis of PI(3)P and concomitant PI(4)P synthesis by PI4K2 α generates distinct recycling endosomal membrane domains destined for exocyst-dependent fusion with the plasma membrane (Fig. 4j), a compartment enriched in PI(4)P^{1,24}. This phosphoinositide conversion mechanism is driven by coincident detection² as well as by reciprocal interactions among its components. These include the association of exocyst and its key regulators with PI4K2 α and PI(4)P²⁰ and with MTM1, facilitation of membrane recruitment of MTM1 by PI4K2 α , as well as association of MTM1 both with PI4K2 α and with exocyst. A similar phosphoinositide conversion mechanism involving MTM1 and/or other myotubularin family members⁶ may underlie cargo exit from endosomes towards PI(4)P- and PI(4,5)P₂-containing lysosomes and autolysosomes. Importantly, our results indicate a close functional correlation between MTM1 depletion and XLCNM-patient mutations, and they suggest that defective phosphoinositide conversion at endosomes may account for the disturbed muscle morphology and impaired β 1-integrin-dependent muscle attachment in developing muscle fibres in patients with XLCNM⁶. The fact that integrin localization defects in cells from patients with XLCNM can be reversed by application of Vps34 inhibitors¹⁵ may offer a potential avenue to combat this devastating disease.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 October; accepted 4 December 2015.

Published online 13 January 2016.

1. Balla, T. Phosphoinositides: tiny lipids with giant impact on cell regulation. *Physiol. Rev.* **93**, 1019–1137 (2013).
2. Di Paolo, G. & De Camilli, P. Phosphoinositides in cell regulation and membrane dynamics. *Nature* **443**, 651–657 (2006).
3. Guo, J. *et al.* Phosphatidylinositol 4-kinase type II α is responsible for the phosphatidylinositol 4-kinase activity associated with synaptic vesicles. *Proc. Natl Acad. Sci. USA* **100**, 3995–4000 (2003).
4. Hammond, G. R., Machner, M. P. & Balla, T. A novel probe for phosphatidylinositol 4-phosphate reveals multiple pools beyond the Golgi. *J. Cell Biol.* **205**, 113–126 (2014).
5. Raiborg, C., Schink, K. O. & Stenmark, H. Class III phosphatidylinositol 3-kinase and its catalytic product PtdIns3P in regulation of endocytic membrane traffic. *FEBS J.* **280**, 2730–2742 (2013).
6. Amoasii, L., Hnia, K. & Laporte, J. Myotubularin phosphoinositide phosphatases in human diseases. *Curr. Top. Microbiol. Immunol.* **362**, 209–233 (2012).

7. Dowling, J. J. *et al.* Loss of myotubularin function results in T-tubule disorganization in zebrafish and human myotubular myopathy. *PLoS Genet.* **5**, e1000372 (2009).
8. Herman, G. E., Finegold, M., Zhao, W., de Gouyon, B. & Metzner, A. Medical complications in long-term survivors with X-linked myotubular myopathy. *J. Pediatr.* **134**, 206–214 (1999).
9. Buj-Bello, A. *et al.* The lipid phosphatase myotubularin is essential for skeletal muscle maintenance but not for myogenesis in mice. *Proc. Natl Acad. Sci. USA* **99**, 15060–15065 (2002).
10. Ribeiro, I., Yuan, L., Tanentzapf, G., Dowling, J. J. & Kiger, A. Phosphoinositide regulation of integrin trafficking required for muscle attachment and maintenance. *PLoS Genet.* **7**, e1001295 (2011).
11. Hsu, V. W., Bai, M. & Li, J. Getting active: protein sorting in endocytic recycling. *Nature Rev. Mol. Cell Biol.* **13**, 323–328 (2012).
12. Cao, C., Backer, J. M., Laporte, J., Bedrick, E. J. & Wandinger-Ness, A. Sequential actions of myotubularin lipid phosphatases regulate endosomal PI(3)P and growth factor receptor trafficking. *Mol. Biol. Cell* **19**, 3334–3346 (2008).
13. Cullen, P. J. & Korswagen, H. C. Sorting nexins provide diversity for retromer-dependent trafficking events. *Nature Cell Biol.* **14**, 29–37 (2012).
14. Munson, M. J. *et al.* mTOR activates the VPS34-UVRAG complex to regulate autolysosomal tubulation and cell survival. *EMBO J.* **34**, 2272–2290 (2015).
15. Bago, R. *et al.* Characterization of VPS34-IN1, a selective inhibitor of Vps34, reveals that the phosphatidylinositol 3-phosphate-binding SGK3 protein kinase is a downstream target of class III phosphoinositide 3-kinase. *Biochem. J.* **463**, 413–427 (2014).
16. Lorenzo, O., Urbé, S. & Clague, M. J. Systematic analysis of myotubularins: heteromeric interactions, subcellular localisation and endosome related functions. *J. Cell Sci.* **119**, 2953–2959 (2006).
17. Subramanian, D. *et al.* Activation of membrane-permeant caged PtdIns(3)P induces endosomal fusion in cells. *Nature Chem. Biol.* **6**, 324–326 (2010).
18. Wang, Y. J. *et al.* Phosphatidylinositol 4 phosphate regulates targeting of clathrin adaptor AP-1 complexes to the Golgi. *Cell* **114**, 299–310 (2003).
19. Minogue, S. *et al.* Phosphatidylinositol 4-kinase is required for endosomal trafficking and degradation of the EGF receptor. *J. Cell Sci.* **119**, 571–581 (2006).
20. Mizuno-Yamasaki, E., Medkova, M., Coleman, J. & Novick, P. Phosphatidylinositol 4-phosphate controls both membrane recruitment and a regulatory switch of the Rab GEF Sec2p. *Dev. Cell* **18**, 828–840 (2010).
21. Takahashi, S. *et al.* Rab11 regulates exocytosis of recycling vesicles at the plasma membrane. *J. Cell Sci.* **125**, 4049–4057 (2012).
22. Hertzog, M. & Chavrier, P. Cell polarity during motile processes: keeping on track with the exocyst complex. *Biochem. J.* **433**, 403–409 (2011).
23. Jović, M. *et al.* Endosomal sorting of VAMP3 is regulated by PI4K2A. *J. Cell Sci.* **127**, 3745–3756 (2014).
24. Hammond, G. R. V. *et al.* PI4P and PI(4,5)P₂ are essential but independent lipid determinants of membrane identity. *Science* **337**, 727–730 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Pluska for initial studies on PI4K2 α -exocyst, I. Ganley for fluorescently labelled p40-Phox, M. Ringling, M. Mühlbauer, L. von Oertzen, and S. Zillmann for technical assistance, and A. Marat for comments. This work was supported by grants from the German Research Foundation (SFB740/C8 to V.H. and SFB958/A11 to M.K.) and Agence Nationale de la Recherche (ANR-14-CE12-0009/13-BSV2-0004 to J.L.).

Author Contributions K.K., D.P., M.K. and A.-S.N. performed experiments; R.M., D.S., and J.L. contributed reagents; K.K., M.W., M.K., C.S., and V.H. designed research. K.K. and V.H. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to V.H. (haucke@fmp-berlin.de).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Antibodies. The following antibodies were used in this study: mouse-anti-actin (clone ac-15, Sigma-Aldrich, western blotting (WB): 1:10,000), mouse-anti-AP-1 γ (γ -adaplin, clone 100/3, Sigma-Aldrich, immunofluorescence (IF): 1:100), mouse-anti-AP-2 α (α -adaplin, clone AP-6, hybridoma cell line, IF: 1:100), rabbit-anti-APPL1 (Cell Signalling, IF: 1:100), mouse-anti-B10 (IGBMC, WB: 1:2,000), mouse-anti- β 1-integrin (clone LM534, Millipore, IF: 1:375), mouse-anti- β 1-tubulin (clone B5-1-2, Sigma-Aldrich, WB: 1:500), rabbit-anti-clathrin heavy chain (Abcam, IF: 1:500), mouse-anti-clathrin heavy chain (clone TD1, hybridoma cell line, WB: 1:500), rabbit-anti-EEA1 (Cell Signalling, IF: 1:100), mouse-anti-EGFR (clone R-1, Santa Cruz, IF: 1:100), mouse-anti-Exo70 (Millipore, WB: 1:500, IF: 1:100), rabbit-anti-Gadkin (ref. 25, WB: 1:1,000), mouse-anti-GM130 (BD Transduction, IF: 1:100), rabbit-anti-GFP (Abcam, WB: 1:10,000, IF: 1:500), mouse-anti-HA (clone HA.11, Covance, IF: 1:400), rabbit-anti-HA (Cayman Chemical, IF: 1:100), rabbit-anti-HA (clone Y-11, Santa Cruz, WB: 1:500), mouse-anti-HA-Alexa Fluor 488 (clone HA.11, Covance, IF: 1:100), mouse-anti-LAMP1 (BD Pharmingen, IF: 1:200), mouse-anti-LC3B (clone 4E12, MBL International, IF: 1:100), rabbit-anti-MTM1 (raised against amino acid 19-33 and amino acid 502-516 of human MTM1, WB: 1:250), mouse-anti-PI(4)P (catalogue: Z-P004, Echelon Biosciences, IF: 1:63), mouse-anti-PI(4,5)P₂ (catalogue: Z-A045, Echelon Biosciences, IF: 1:200), mouse-anti-PI(3,4)P₂ (catalogue: Z-P034b, Echelon Biosciences, IF: 1:150), rabbit-anti-PI4K2 α (ref. 26, WB: 1:2,000), sheep-anti-PIKfyve (Tocris Biosciences, WB: 1:1,000), mouse-anti-Rab5 (BD Transduction, IF: 100), rabbit-anti-Rab7 (clone D95F2, Cell Signalling, IF: 1:50), rabbit-anti-Rab11a (Life Technologies, WB: 1:500), rabbit-anti-Sec3 (Proteintech Group, WB: 1:500), mouse-anti-Sec6 (Stressgen, WB: 1:500), mouse-anti-Sec8 (BD Transduction, WB: 1:500), mouse-anti-SNX4 (Sigma-Aldrich, WB: 1:500), mouse-anti-SNX17 (Proteintech Group, WB: 1:1,000), sheep-anti-TGN46 (Serotec, IF: 1:200), mouse-anti-TfR (clone H68.4, Life Technologies, IF/flow cytometry: 1:200), rabbit-anti-TfR (Sigma-Aldrich, IF: 1:100), rabbit-anti-Vps26 (Abcam, IF: 1:100), rabbit-anti-Vps34 (clone D9A5, Cell Signalling, WB: 1:1,000).

siRNAs. All siRNAs used in this study were 21-, 23-, or 27-base oligonucleotides including 3'-dTdT overhangs. For silencing, the following siRNAs were used targeting the human isoform: Exo70 5'-GGTTAAAGGTGACTGATTA-3', MTM1 5'-GTGCAAGACCGAGCGTAA-3', MTMR1 5'-GAGATAGTGTGCAAGGATA-3', MTMR2 5'-GGACATCGATTTCAACTAA-3', MTMR4 5'-CATAGGTTACGGCAAA-3', MTMR7 5'-TGCAAGAACTTTCAGATAA-3', PI4K2 α 5'-GGATCATTGCTGCTCTTCAA-3', Rab11a 5'-AAGAGCGATATCGAGCTATAA-3', Sec3 5'-CCTGTTGGATATGGGAAACAT-3', Sec6 5'-CTGGAGGCGAGCATCAACAC-3', SNX4 5'-TGGTTCAGAGTGTCTTAACA-3', SNX17 5'-CTGGCTTTTGAATACCTCA-3', and Vps34 5'-CCCATGAGATGTACTTGAACGTAAT-3'. For silencing Kif16b and PIKfyve, a pool of four siRNAs was obtained from Dharmaco (Thermo Scientific). The scrambled control siRNA used throughout this study corresponded to the scrambled γ 1 adaplin sequence 5'-AAATCGGATATCGGAATAG-3'.

Plasmids. Complementary DNA encoding full-length human MTM1 and MTMR2 was provided by G. Di Paolo and inserted into a pcDNA3.1(+)-based haemagglutinin (HA)-, mCherry-, or eGFP-expression vector with tags at the amino (N) terminus of MTM1 and MTMR2. siRNA-resistant MTM1 constructs were created by introducing four silent mutations: 5'-gatccaggcctagtgttaa-3'. P205L, R241L, Y397C, and C375S mutants of MTM1 were created by mutation of the respective amino acid of human MTM1. N-terminally tagged human MTMR7 and mouse MTMR1 were in a pcDNA3.1(+)-based vector backbone. N-terminally tagged GFP-MTMR4 was a gift from M. Clague. N-terminally tagged GFP-Rab4A and GFP-Rab11A were a gift from P. van Sluis, and GFP-Rab14 was provided by T. Proikas-Cezanne. N-terminally tagged GFP-Rab5A GFP-Rab8A and GFP-Rab35 were in the pEGFP-C vector backbone (Clontech) and Rab5A in a pcDNA3.1(+)-vector backbone. Q79L mutant of Rab5A was created by mutation of the respective amino acid of human Rab5A. Full-length human SNX1, SNX3, SNX4, SNX8, SNX15, SNX17, and SNX27 were inserted in a pcDNA3.1(+)-based expression vector to express N-terminally tagged eGFP-fusion proteins except for carboxy (C)-terminally tagged SNX17-eGFP. N-terminally tagged eGFP-2xFYVE-Hrs, eGFP-2xPH-PLC δ and eGFP-2xPH-FAPP1 were in a pcDNA3.1(+)-based expression vector. The mRFP-2xPH-TAPP1 was a gift from T. Takenawa and the eGFP-2xPH domain of Bruton's tyrosine kinase was a gift from M. Wymann. N-terminally tagged eGFP-mPI4K2 α was in a pcDNA3.1(+)-based expression vector. N-terminally tagged HA-mPI4K2 α was in the pcDNA5/FRT/TO expression vector (Life Technologies). D308A mutant of PI4K2 α was created by mutation of the respective amino acid of mouse PI4K2 α . To create N-terminally tagged

GST-MTM1 and -Sec6, full-length human MTM1 and Sec6 were cloned in the Gateway pENTRA1 entry vector and by recombination transferred into the Gateway pGEX4T3 vector (Life Technologies). N-terminally tagged mouse PI4K2 α was in the pGEX4T1 vector backbone. The B10-tag was engineered in-house by inserting the B10 epitope of the human oestrogen receptor in the pSG5 vector backbone (Stratagene) and used to create N-terminally tagged B10-MTM1 and B10-Sec6 (pSG5HERB10 empty vector).

Cell lines. HeLa and COS-1 cells were from ATCC and not used beyond passage 30 from original derivation from ATCC without further authentication. Hek293 cells stably expressing HA-tagged mPI4K2 α were generated using the FlpIn system developed by Life Technologies according to the manufacturer's protocol. The fibroblast cell line H31 from patients with XLCNM has a genomic deletion of the entire MTM1 (refs 27, 28); the fibroblast cell line G92-628 from a patient with XLCNM, referred to as XLCNM patient 2, has a stop mutation in MTM1 at amino acid 37 (refs 27, 28). HDFa cells (human dermal fibroblast from adult healthy individuals) were obtained from Life Technologies. H31, G92-628, and HDFa cells were cultured in MEM (Life Technologies) containing 15% FCS and not used beyond passage 20. HeLa-M Clone 1 (C1) cells were cultured in DMEM (4.5 g l⁻¹ glucose, Life Technologies) containing 10% FCS and 1.66 μ g ml⁻¹ puromycin and not used beyond passage 10 (ref. 29). All cell lines were routinely tested for mycoplasma contaminations on a monthly basis.

siRNA and plasmid transfections. HeLa cells were transfected with siRNA using Oligofectamin (Life Technologies) according to the manufacturer's protocol. To achieve optimal knockdown efficiency, two rounds of silencing were performed. Cells were transfected on day 1, expanded on day 2, transfected for a second time on day 3, seeded for the experiment on day 4, and the experiment was performed on day 5. For transient overexpression of proteins in knockdown cells, plasmids were transfected on day 4 using Lipofectamin 2000 (Life Technologies) according to the manufacturer's protocol. In case of co-transfection of two plasmids, JetPrime (Polyplus) was used according to the manufacturer's protocol.

For transient overexpression of proteins in untreated cells, plasmids were transfected 24 h before analysis using Lipofectamin 2000 (Life Technologies).

For knockdown of transiently overexpressed proteins, cells were simultaneously transfected with plasmids and siRNA using Lipofectamin 2000 according to the manufacturer's protocol; expression was allowed overnight and cells analysed the next day.

HDFa, H31, and COS-1 cells were transfected using Lipofectamin 3000 (Life Technologies) according to the manufacturer's protocol; expression was allowed overnight and cells analysed the next day.

Transferrin uptake and surface labelling. Cells seeded on coverslips coated with Matrigel (BD Biosciences) were serum-starved for 2 h and used for either Tf uptake or surface labelling.

For quantitative Tf uptake, cells were treated with 25 μ g ml⁻¹ Tf-Alexa647 (Life Technologies) for 10 min at 37°C. After being washed twice with ice-cold phosphate-buffered saline (PBS), cells were acid washed at pH 5.3 (0.2 M Na-acetate, 0.2 M NaCl) for 1 min on ice to remove surface-bound Tf, followed by washing twice with ice-cold PBS and fixation with 4% paraformaldehyde (PFA) for 45 min at room temperature (23–25°C). For all immunocytochemistry stainings except MTM1 or Rab5 Q79L-expressing cells, inhibitor treatments (Wortmannin or Vps34-IN1), and staining of β 1-integrin, cells were treated with 25 μ g ml⁻¹ Tf-Alexa647 (Life Technologies) for 30 min at 37°C to allow Tf uptake to reach saturation. Cells were washed twice with PBS, followed by immunocytochemistry staining as described in Immunocytochemistry and spinning disc confocal imaging.

For TfR surface labelling, cells were incubated with 25 μ g ml⁻¹ Tf-Alexa647 for 45 min at 4°C to block endocytosis, washed three times with ice cold PBS, and fixed with 4% PFA for 45 min at room temperature.

Tf uptake and surface labelling, and EGFR and LC3B labelling, were analysed using a Nikon Eclipse Ti microscope (eGFP filter set: F36-526; TexasRed filter set: F36-504; Cy5 filter set: F46-009; DAPI filter set: F46-000), equipped with a \times 40 oil-immersion objective (Nikon), a sCMOS camera (Neo, Andor), and a 200 W mercury lamp (Lumen 200, Prior), operated by open-source ImageJ-based micro-manager software and quantified using open-source ImageJ software. Levels of eGFR, LC3B, and internalized or surface-bound Tf were normalized to the cell area. The ratio of internalized Tf to surface-bound Tf was used to distinguish between uptake and recycling defects. Owing to small differences in TfR level in HDFa and H31 cells, surface-bound Tf was normalized to TfR level. TfR levels in HDFa and H31 cells were analysed by TfR antibody staining (see Immunocytochemistry and spinning disc confocal imaging) and analysed using a Nikon Eclipse Ti microscope as described above.

Immunocytochemistry and spinning disc confocal imaging. Cultured cells seeded on Matrigel-coated coverslips were fixed for 10 min with 4% PFA, washed

twice with PBS, permeabilized in blocking solution (10% goat serum, 20 mM $\text{Na}_2\text{H}_2\text{PO}_4$ pH 7.4, 0.3% Triton X-100, 100 mM sodium chloride) for 30 min, and incubated with primary antibodies diluted in blocking solution for 1 h. After three washes with washing solution (20 mM $\text{Na}_2\text{H}_2\text{PO}_4$ pH 7.4, 0.3% Triton X-100, 100 mM NaCl), secondary antibodies diluted in blocking solution were incubated for 1 h, followed by three washes in washing solution.

For transient overexpression of eGFP-MTM1, cells were washed twice with ice-cold PBS, incubated with PEM (80 mM Pipes pH 6.8, 5 mM EGTA, 1 mM MgCl_2) containing 0.05% saponin for 5 min at 4 °C, followed by a brief wash with ice-cold PEM. After fixation with 3% PFA for 15 min at 4 °C, cells were washed three times with PBS at room temperature, incubated with 50 mM NH_4Cl for 10 min and washed again twice with PBS. Immunocytochemistry was done as described above with the exception that Triton X-100 was exchanged for 0.05% saponin in all buffers.

PI(3)P and PI(4)P stainings were performed as previously described³⁰, although at a reduced concentration of eGFP-2xFYVE of $0.25 \mu\text{g ml}^{-1}$. If indicated, PI(3)P was labelled using a GST-tagged Phox domain of p40 chemically conjugated to Alexa Fluor 488, which was a gift from I. Ganley. For wortmannin (Sigma-Aldrich) treatment, cells were incubated with $2 \mu\text{M}$ wortmannin (dissolved in dry dimethylsulfoxide (DMSO)) or DMSO, diluted in serum-free medium, for 30 min at 37 °C, and subsequent PI(3)P staining was performed as described above. PI(3,4)P₂ and PI(4,5)P₂ stainings were performed as previously described³¹.

For all LC3 immunocytochemistry stainings, fresh serum-containing medium was added 2 h before fixation. In case of bafilomycin A1 (BafA1, Sigma-Aldrich) treatment cells were washed three times with HBSS and incubated with 100 nM BafA1 (dissolved in dry DMSO) diluted in HBSS for 3 h. Control cells were incubated with DMSO diluted in serum-containing medium. After fixation with 4% PFA for 30 min, cells were washed three times with PBS, permeabilized with $200 \mu\text{g ml}^{-1}$ digitonin diluted in PBS for 10 min, followed by three washes with PBS. Cells were incubated with the primary antibody diluted in PBS for 1 h, followed by three washes with PBS. Secondary antibodies diluted in PBS were incubated for 1 h, followed by three washes in PBS.

Protein and lipid immunocytochemistry stainings were routinely analysed and quantified using a spinning disc confocal microscope (Ultraview ERS, Perkin Elmer) with Velocity imaging software (Improvision, Perkin Elmer). For all quantifications, protein and lipid stainings were normalized to cell area. For quantification of TfR localization, cells with either perinuclear or peripheral TfR localization were counted. Peripheral TfR localization was defined as cells with either TfR dispersion or TfR accumulations at the cell periphery as shown in example images (Fig. 2c), and the normalized fraction of cells with perinuclear TfR localization was quantified. The amount of co-localization between two channels was quantified using thresholded Pearson's correlation coefficients. To quantify the amount of co-localization at peripheral sites, thresholded Pearson's coefficients were calculated in three randomly chosen $100 \text{ pixel} \times 100 \text{ pixel}$ squares in the cell periphery. For averaged line scans, line profiles were calculated as the mean fluorescence intensity averaged over 100 pixels. Maximum intensity projections were calculated from z-stacks with 200 nm spacing between slices covering the whole cell. For statistical analysis, see Statistical analysis of immunocytochemistry and live-cell TIRF experiments.

Cholera toxin uptake. Cells seeded on Matrigel-coated coverslips were serum-starved for 30 min and treated with $1 \mu\text{g ml}^{-1}$ cholera toxin subunit B (Ctx) CF568 (Biotium) for 45 min at 37 °C, followed by 30 min chase in starvation medium. After being washed once with PBS, cells were fixed with 4% PFA for 30 min, washed twice with PBS, permeabilized in blocking solution (10% goat serum, 20 mM $\text{Na}_2\text{H}_2\text{PO}_4$ pH 7.4, 0.3% Triton X-100, 100 mM sodium chloride) for 15 min and incubated with primary antibody diluted in blocking solution for 30 min. After three brief washes with PBS, secondary antibody diluted in blocking solution was incubated for 30 min, followed by three brief washes in PBS.

Secretion assay. C1 cells seeded on Matrigel-coated coverslips were washed once with PBS and treated with $1 \mu\text{M}$ D/D solubilizer (Clontech) diluted in HBSS to initiate secretion of the reporter construct. To halt secretion at the indicated time points, cells were placed on ice, washed twice with ice-cold PBS containing 10 mM MgCl_2 and fixed with 4% PFA for 20 min at room temperature, followed by washing twice with PBS. Secretion of the GFP-tagged reporter construct was analysed using a Nikon Eclipse Ti microscope (see Transferrin uptake and surface labelling). Fifteen minutes after initiating secretion the reporter is localized at the Golgi complex. To calculate the secretion from the Golgi, all time points were normalized to 15 min.

TIRF microscopy. TIRF microscopy was performed using a Nikon Eclipse Ti microscope, equipped with an incubation chamber (37 °C), a $\times 60$ TIRF objective (oil-immersion, Nikon), a sCMOS camera (Neo, Andor), a 200 W mercury lamp (Lumen 200, Prior), a triple-colour TIRF setup (laser lines: 488 nm, 568 nm,

647 nm), and operated by open-source ImageJ-based micromanager software. Cells seeded on Matrigel-coated coverslips were treated with $50 \mu\text{g ml}^{-1}$ Tf-Alexa647 (Life Technologies), diluted in serum-free medium for 30 min at 37 °C and 5% CO_2 . For analysis of transferrin exocytosis, time-lapse movies of 15–30 s with a frame rate of 5 Hz were recorded. For all time-lapse movies, the 488 nm channel was acquired before the 647 nm channel, except for Extended Data Fig. 9a, g. Fusion events with the plasma membrane were defined by their characteristic time course (appearance, broadening/spreading of the fluorescence signal, and disappearance), counted and normalized to cell area. Representative kymographs were chosen over 30 s along a line of 400 pixels in length. For statistical analysis, see Statistical analysis of immunocytochemistry and live-cell TIRF experiments.

PIP/AM experiments. Cell-permeable acetoxy methylester (AM)-protected phosphatidylinositol derivatives were synthesized according to published procedures¹⁷. For treatment of cells, PI(3)P/AM and PI(4)P/AM were dissolved in dry DMSO and mixed with an equal volume of 10% pluronic F127 in DMSO (Sigma-Aldrich). PIP/AMs were diluted in serum-free medium to a final concentration of $100 \mu\text{M}$. For immunocytochemistry stainings, cells seeded on Matrigel-coated coverslips were treated with DMSO + pluronic F127 (control) or PIP/AMs for 30 min at 37 °C and then processed as described above. For TfR immunocytochemistry stainings, cells were stimulated with $25 \mu\text{g ml}^{-1}$ Tf-Alexa647 during the PIP/AM treatment. For live-cell TIRF imaging, cells seeded on Matrigel-coated coverslips were treated with $25 \mu\text{g ml}^{-1}$ Tf-Alexa647 and DMSO + pluronic F127 (control) or PIP/AMs for 30 min at 37 °C, directly followed by live-cell imaging.

Vps34-IN1 experiments. For the analysis of $\beta 1$ -integrin accumulations, H31 or HDFA cells were seeded on Matrigel-coated coverslips and treated with DMSO or Vps34-IN1 at a concentration of $0.01\text{--}1 \mu\text{M}$ dissolved in dry DMSO and diluted in serum-containing medium for 48 h, adding freshly diluted Vps34-IN1 or DMSO after 24 h. $\beta 1$ -integrin staining was performed as described in Immunocytochemistry and spinning disc confocal imaging.

For live-cell TIRF imaging (see Fig. 3d), cells seeded on Matrigel-coated coverslips were treated with $50 \mu\text{g ml}^{-1}$ Tf-Alexa647 and DMSO or $1 \mu\text{M}$ VPS34-IN1 for 60 min at 37 °C, directly followed by live-cell imaging in the presence of DMSO or VPS34-IN1, respectively.

¹²⁵I-labelled EGF degradation. Human EGF was purchased from Peprotech and ¹²⁵I-labelled EGF from Perkin Elmer. ¹²⁵I-labelled EGF degradation was performed as described previously³².

¹²⁵I-labelled transferrin uptake and surface labelling. Human holo-transferrin was purchased from Sigma-Aldrich and ¹²⁵I-labelled Tf from Perkin Elmer. HeLa cells seeded in 24-well plates were starved for 1 h in serum-free medium containing 0.1% BSA and 20 mM HEPES pH 7.4. Cells were stimulated with $1 \mu\text{g ml}^{-1}$ ¹²⁵I-labelled Tf in starvation medium at 37 °C for the indicated time points and washed twice on ice with PBS. Surface-bound ¹²⁵I-labelled Tf was removed by an acid wash with 0.2 M acetic acid, 0.5 M NaCl for 5 min on ice, collected, and radioactivity was measured using a scintillation counter (HIDEX 300SL). Cells were dried at room temperature for 5 min, lysed with 1 M NaOH for 60 min and the radioactivity of the lysate measured (corresponding to internalized Tf). Non-specific binding was measured for each time-point in the presence of a 300-fold excess of cold Tf and was subtracted from all values. The ratio of internalized to surface-bound Tf was plotted over time and Michaelis constant (K_m) values were calculated using Prism software (GraphPad).

Flow cytometry. On ice, cells were washed once with ice-cold PBS and detached from the culture dish by incubating for 5 min on ice with 0.1% PronaseE (Sigma-Aldrich), 0.5 mM EDTA solution in PBS. Cells were resuspended in PBS, pelleted at 300 g for 5 min at 4 °C and fixed in 4% PFA, 4% sucrose in PBS for 20 min at room temperature. Ten times excess volume of blocking solution (0.05% Saponin, 0.01% BSA in PBS) was added, then cells were pelleted and resuspended in blocking solution. After 15 min primary antibody (mouse-anti-TfR) was added and incubated for an additional 1 h. After washing once with blocking solution, cells were incubated for 1 h with secondary antibody diluted in blocking solution, followed by washing once with 0.2% BSA in PBS. Cells were resuspended in 0.2% BSA in PBS and analysed by flow cytometry using a BD LSRFortessa.

Membrane fractionation. Cells from a 100% confluent 10 cm cell culture dish were harvested in homogenization buffer (20 mM HEPES pH 7.4, 100 mM KCl, 2 mM MgCl_2 , 1 mM PMSF, 0.1% protease inhibitor cocktail) and homogenized using a European Molecular Biology Laboratory cell cracker (HGM; inner diameter 8.020 mm, ball diameter 8.004 mm, with 12 strokes), followed by three freeze-thaw cycles in liquid nitrogen. Total cell lysate was collected after centrifugation at $1,000 \text{ g}$ for 5 min at 4 °C. To obtain the cytosolic fraction, total cell lysate was centrifuged at $100,000 \text{ g}$ for 30 min at 4 °C, the supernatant was collected, and protein concentration and volume determined. The membrane pellet was washed once in homogenization buffer and collected in a volume corresponding to the volume of the cytosol fraction. Equal volumes of total cell lysate (minimal concentration

0.125 $\mu\text{g}\mu\text{l}^{-1}$, corresponding to minimal loading of 5 μg), membrane pellet, and cytosol fraction were loaded onto a 10% acrylamide gel for SDS–polyacrylamide gel electrophoresis (SDS–PAGE) followed by immunoblotting. Western blot development was done using a LI-COR Odyssey Fc imager, and western blot bands were quantified using Image Studio Lite Version 4.0 software (LI-COR). For Exo70, Sec3, Sec8, and MTM1, protein levels in the total cell lysate were normalized to actin, whereas protein levels in the membrane fraction were normalized to gadkin. The ratio of membrane to total cell lysate was used to quantify the membrane fraction of Exo70, Sec3, Sec8, and MTM1. All knockdown conditions were normalized to membrane fractions of the respective protein in scrambled siRNA-treated controls. Data are presented as mean values \pm s.e.m. from five independent experiments (n). Statistical testing was performed using a one-sample t -test.

Co-immunoprecipitation experiments. Expression of HA-tagged PI4K2 α in stably transfected Hek FlpIn cells was induced overnight by addition of doxycycline. As a control, untransfected Hek FlpIn cells were used. Cells were harvested in lysis buffer (20 mM HEPES pH 7.4, 100 mM KCl, 2 mM MgCl_2 , 1 mM PMSE, 0.1% protease inhibitor cocktail, 1% Triton X-100) and incubated on ice for 30 min, followed by centrifugation at 43,500 g for 20 min at 4°C. The supernatant was centrifuged again at 265,000 g for 15 min at 4°C. HA-matrix beads (Covance, mouse-anti-HA.11) were used to immunoprecipitate HA-tagged PI4K2 α . Protein (3–5 mgml^{-1}) was loaded on 30 μl 1:1 washed HA-matrix beads slurry and incubated for 1 h at 4°C on a rotating wheel. Beads were pelleted, washed twice with lysis buffer, followed by two washes with homogenization buffer, and bound protein was eluted in 60 μl 1 \times Laemmli sample buffer. Eluates were loaded onto a 10% acrylamide gel for SDS–PAGE followed by immunoblotting.

Affinity chromatography/GST-pulldown assays. Pulldown assays using GST-tagged PI4K2 α as a bait were performed as described previously²⁶.

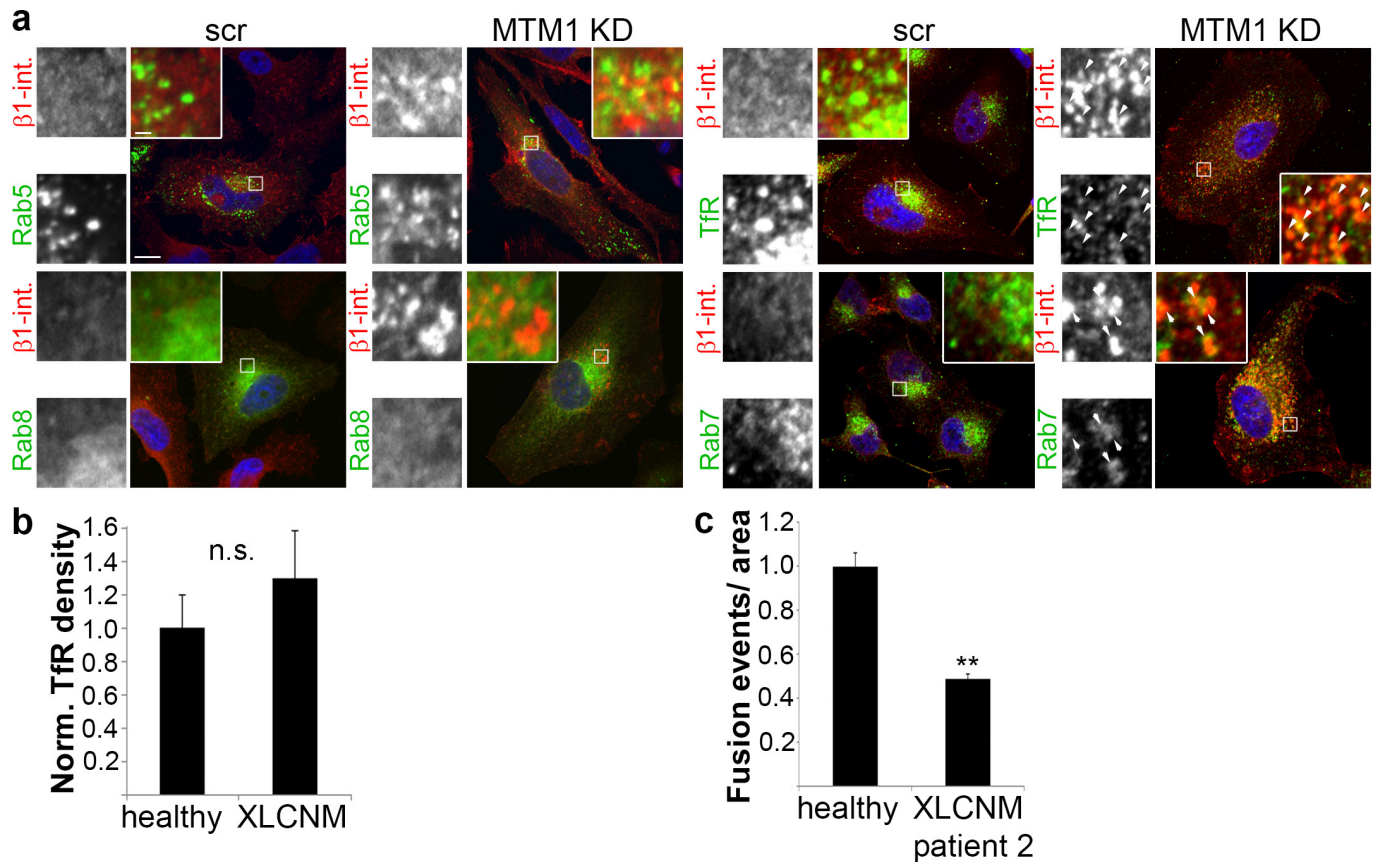
For pulldown assays using GST-tagged MTM1 and Sec6, GST fusion proteins were expressed in *Escherichia coli* BL21 pRARE strain. For negative control, the empty pGEX4T3 vector (GST alone) was used. The induction of expression was performed with 1 mM IPTG at 16°C for 14 h. Bacteria were lysed on ice by sonication in lysis buffer (50 mM Tris HCl pH7, 200 mM NaCl, 1 mM EDTA, 1 mM DTT, complete protease inhibitor cocktail (PIC, Roche)) supplemented with 1 mgml^{-1} lysozyme, 1 mM PMSE, 0.1% sarcosyl, 0.5% Triton X-100, and then centrifuged at 15,000 g for 30 min. GST-tagged proteins were purified from bacterial lysates by incubation with Glutathione Sepharose 4B beads (GE Healthcare) for 1 h followed by extensive washing with lysis buffer plus 1 mM PMSE. In parallel, COS-1 cells expressing B10-MTM1, B10-SEC6, HA-PI4K2 α , or corresponding controls were lysed with a buffer containing 20 mM HEPES-KOH pH 7.5, 200 mM NaCl, 0.5% Triton X-100, 10% glycerol, 4 mM DTT, 1 mM EDTA, and PIC by passage through 25-gauge needles. Insoluble material was removed by centrifugation at 11,000 g for 10 min. Twenty micrograms of the purified GST fusion proteins coupled to glutathione beads were then incubated with 350 μg of COS-1 cell extracts. After washing beads three times with a buffer containing 20 mM HEPES-KOH, 200 mM NaCl, 1 mM DTT, and PIC (pH 7.3), 10 μg of GST beads were analysed by electrophoresis on a 10% SDS–polyacrylamide gel. Bound B10-tagged MTM1 and Sec6 or HA-tagged PI4K2 α were detected with a B10-specific antibody or a PI4K2 α -specific antibody, respectively. The entire procedure was performed at 4°C, unless specified.

Liposome flotation assay. Liposome flotation assays were performed in Hek293 cells as previously described³¹.

Electron microscopy. Scrambled or MTM1 siRNA transfection was performed as described in HeLa cells that were grown in 6 cm plastic dishes. Cells were serum-starved for 2 h and treated with 25 $\mu\text{g}\text{ml}^{-1}$ transferrin-HRP (Fitzgerald) for 30 min at 37°C. Cells were fixed with 2% glutaraldehyde in PBS. After rinsing in fresh PBS, cells were mechanically detached by scratching, pelleted, and embedded into gelatin. After osmification with aqueous 1% osmium tetroxide, samples were stained *en bloc* with 1% aqueous uranyl acetate and embedded in epoxy resin. Sections were viewed with Zeiss 910 transmission electron microscope and micrographs were taken along the cell perimeter at $\times 20,000$. For morphometric analysis, images were combined to reconstruct the perimeter of a cell. The size and number of transferrin-HRP-labelled organelles up to 1 μm distance from the plasma membrane were quantified. The number of Tf-HRP endosomes was normalized to the plasma membrane perimeter. Ten cells per condition were analysed. To calculate the size of Tf-HRP endosomes, 52 endosomes for scrambled siRNA-treated controls and 90 endosomes for MTM1-depleted cells from a total of 10 cells were analysed. Data are presented as mean values \pm s.e.m. A statistical test was performed using a one-sample t -test.

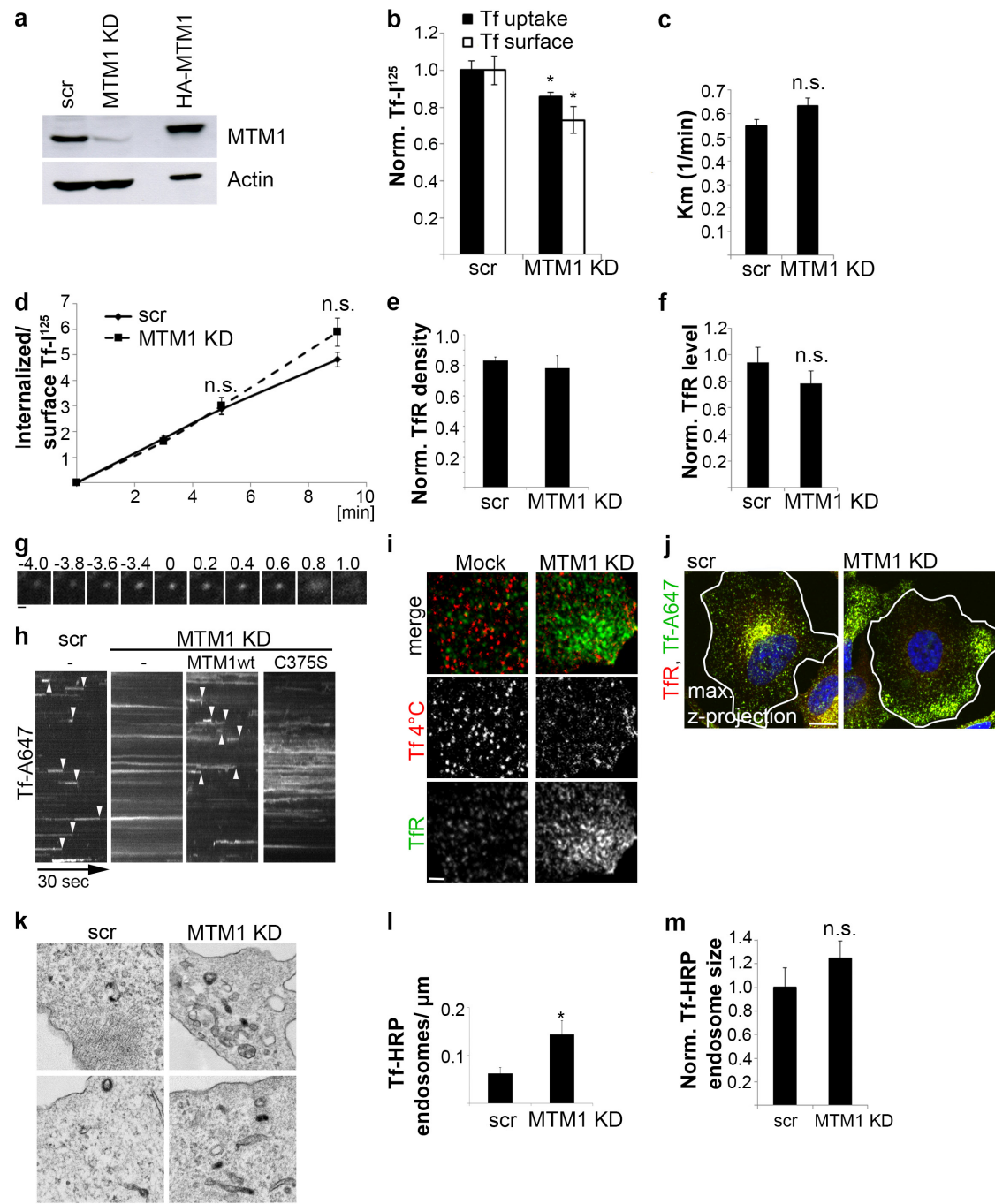
Statistical analysis of immunocytochemistry and live-cell TIRF experiments. For analysis of immunocytochemistry experiments, a minimum of three independent experiments (n) was performed and statistically significant estimates for each sample were obtained by choosing an appropriate sample size, correlating to 15–30 images per condition per experiment for microscopy-based quantifications. Cells were chosen arbitrarily according to the fluorescent signal in a separate channel, which was not used for quantification. Data are presented as mean values \pm s.e.m. For analysis of exocytic events per unit area in live-cell TIRF imaging experiments, a minimum of ten videos (duplicate coverslips: five videos per coverslip) per condition per experiment were acquired and fusion events were counted in minimally five videos per condition per experiment. A minimum of three independent experiments were performed and data represent mean values \pm s.e.m. All statistical tests were performed using a two-tailed, unpaired t -test, without excluding samples from statistical analysis.

25. Schmidt, M. R. *et al.* Regulation of endosomal membrane traffic by a Gadkin/AP-1/kinesin KIF5 complex. *Proc. Natl Acad. Sci. USA* **106**, 15344–15349 (2009).
26. Mössinger, J. *et al.* Phosphatidylinositol 4-kinase II α function at endosomes is regulated by the ubiquitin ligase Itch. *EMBO Rep.* **13**, 1087–1094 (2012).
27. Laporte, J. *et al.* MTM1 mutations in X-linked myotubular myopathy. *Hum. Mutat.* **15**, 393–409 (2000).
28. Laporte, J., Kress, W. & Mandel, J. L. Diagnosis of X-linked myotubular myopathy by detection of myotubularin. *Ann. Neurol.* **50**, 42–46 (2001).
29. Gordon, D. E., Bond, L. M., Sahlender, D. A. & Peden, A. A. A targeted siRNA screen to identify SNAREs required for constitutive secretion in mammalian cells. *Traffic* **11**, 1191–1204 (2010).
30. Hammond, G. R., Schiavo, G. & Irvine, R. F. Immunocytochemical techniques reveal multiple, distinct cellular pools of PtdIns4P and PtdIns(4,5)P₂. *Biochem. J.* **422**, 23–35 (2009).
31. Posor, Y. *et al.* Spatiotemporal control of endocytosis by phosphatidylinositol-3,4-bisphosphate. *Nature* **499**, 233–237 (2013).
32. Diesenberg, K., Beerbaum, M., Fink, U., Schmieder, P. & Krauss, M. SEPT9 negatively regulates ubiquitin-dependent downregulation of EGFR. *J. Cell Sci.* **128**, 397–407 (2015).



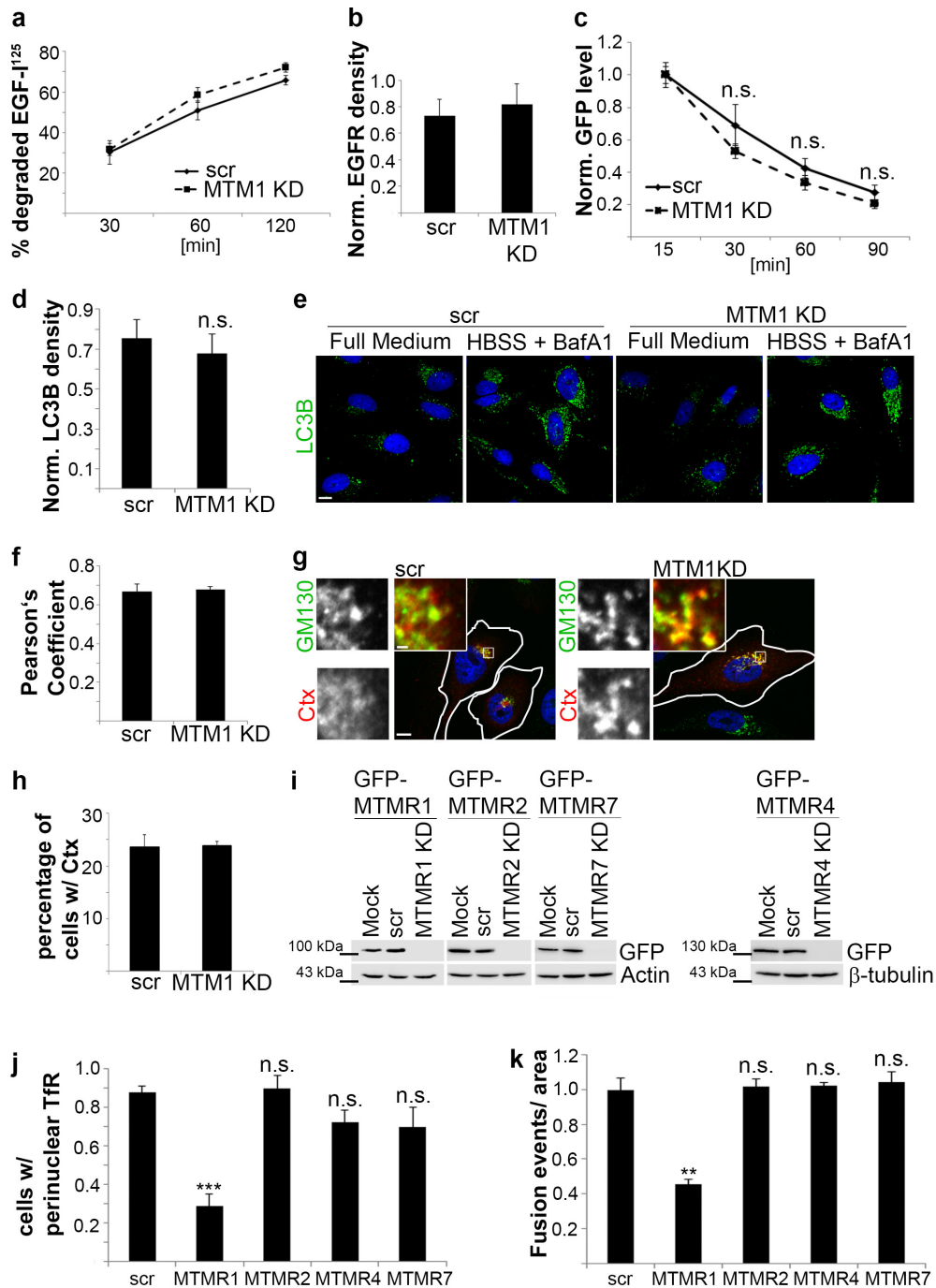
Extended Data Figure 1 | Characterization of XLCNM patient phenotype. **a**, HeLa cells treated with scrambled (scr) siRNA or depleted of MTM1 (MTM1-KD) expressing eGFP-Rab (Rab5, Rab8) or labelled for the indicated protein were co-labelled using β 1-integrin-specific antibodies. Co-localization of β 1-integrin with late endosomal Rab7 and recycling endosomal TfR in MTM1-KD cells is indicated by arrowheads. Scale bar, 10 μ m; magnified insets, 1 μ m. Representative images from one of two independent experiments (for TfR) or from one experiment with

10–20 images per condition are shown. **b**, Quantitative analysis of TfR levels per cell area of XLCNM H31 cells and healthy controls ($n = 3$). **c**, Impaired Tf exocytosis from endosomes in XLCNM G92-628 cells (patient 2) monitored by TIRF microscopy. Quantified are Tf exocytic events normalized to cell area. n , number of independent experiments with 15–30 images (**b**) or 5 videos (**c**) analysed per condition per experiment. Mean \pm s.e.m.; $**P < 0.01$, unpaired two-tailed t -test; NS, non-significant.



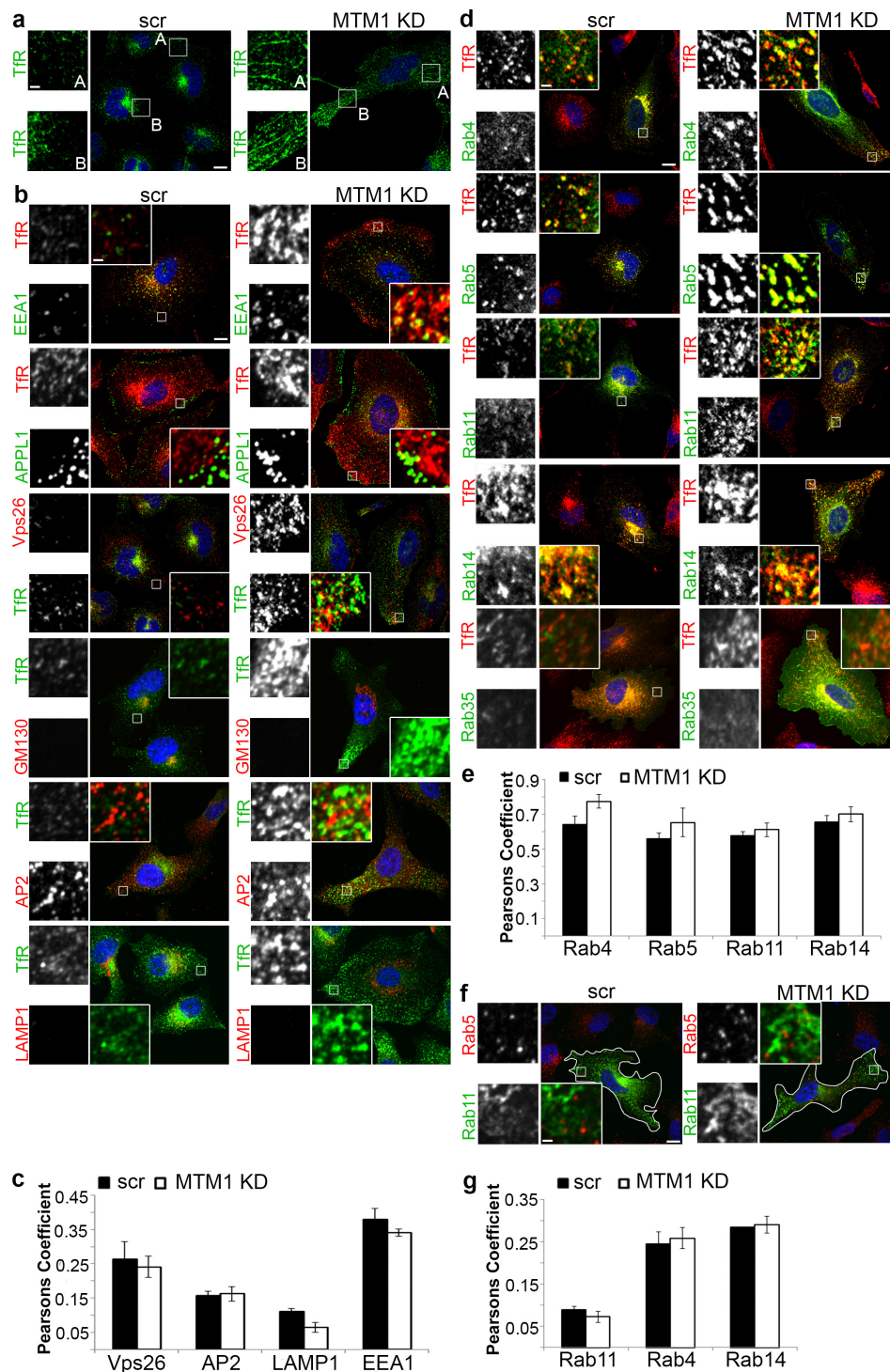
Extended Data Figure 2 | Transferrin endosomes in MTM1-depleted cells. **a**, Efficient siRNA-mediated depletion of MTM1 from HeLa cells as demonstrated by immunoblot analysis. Lysates of HeLa cells expressing HA-MTM1 were taken as a control to assure MTM1 antibody specificity. For blot source data, see Supplementary Fig. 1. **b–d**, Time course of Tf internalization was measured radioactively and the ratio between Tf uptake and surface Tf levels over time was quantified in HeLa cells depleted of MTM1 (MTM1-KD) compared with scrambled (scr) siRNA-treated controls. While Tf endocytic rates remained unchanged in MTM1-KD cells (**c**, **d**), Tf uptake and surface level were significantly decreased in MTM1-depleted cells (**b**), indicative of defective Tf recycling upon loss of MTM1. **b–d**, Mean \pm s.e.m., $n = 4$ independent experiments, each experiment done in technical triplicates; NS, non-significant, $*P < 0.05$, unpaired, two tailed t -test. **e**, **f**, Quantitative analysis of TfR levels by confocal imaging (TfR levels area; **e**) or flow cytometry (TfR levels per cell; **f**) of (scr) control or MTM1-KD cells (mean \pm s.e.m., $n = 3$ independent experiments; NS, non-significant, unpaired two-tailed t -test). **g**, **h**, Impaired Tf exocytosis monitored by TIRF microscopy in MTM1-depleted HeLa cells is restored by re-expression of eGFP-MTM1

WT, but not inactive eGFP-MTM1 (C375S) or eGFP. **g**, Snapshot of a single exocytic event in control cells with characteristic time course of appearance, brightening, and spreading of the fluorescent signal. Scale bar, 400 nm. **h**, Kymographs of Tf fluorescence signal over 28 μ m. Arrowheads indicate time of plasma membrane fusion. **i**, Surface TfR labelled with Tf-Alexa647 (at 4 $^{\circ}$ C) does not co-localize with intracellular TfR in control or MTM1-depleted HeLa cells. Scale bar, 2 μ m. **j**, Maximum intensity projections of z-stacks acquired by confocal imaging confirm the perinuclear accumulation of Tf-ligand and TfR in controls and accumulation at the cell periphery in MTM1-depleted cells as shown in Fig. 2c. Scale bar, 5 μ m. **k**, Electron micrographs of MTM1-depleted HeLa cells treated with Tf-HRP. Scale bar, 200 nm. **l**, Accumulation of Tf-HRP positive endosomes in MTM1-depleted cells (mean \pm s.e.m., $n = 10$ cells, $*P < 0.05$, one-sample t -test), **m**, while size of TfR-HRP positive endosomes remained unchanged (mean \pm s.e.m., $n = 52$ (scr) and 90 (MTM1-KD) endosomes from 10 cells; NS, non-significant, one-sample t -test). **g–k**, Representative images from one of three (**g**, **h**) or two (**i**) independent experiments or from one experiment with ten images (**j**) or ten cells (**k**) per condition are shown.



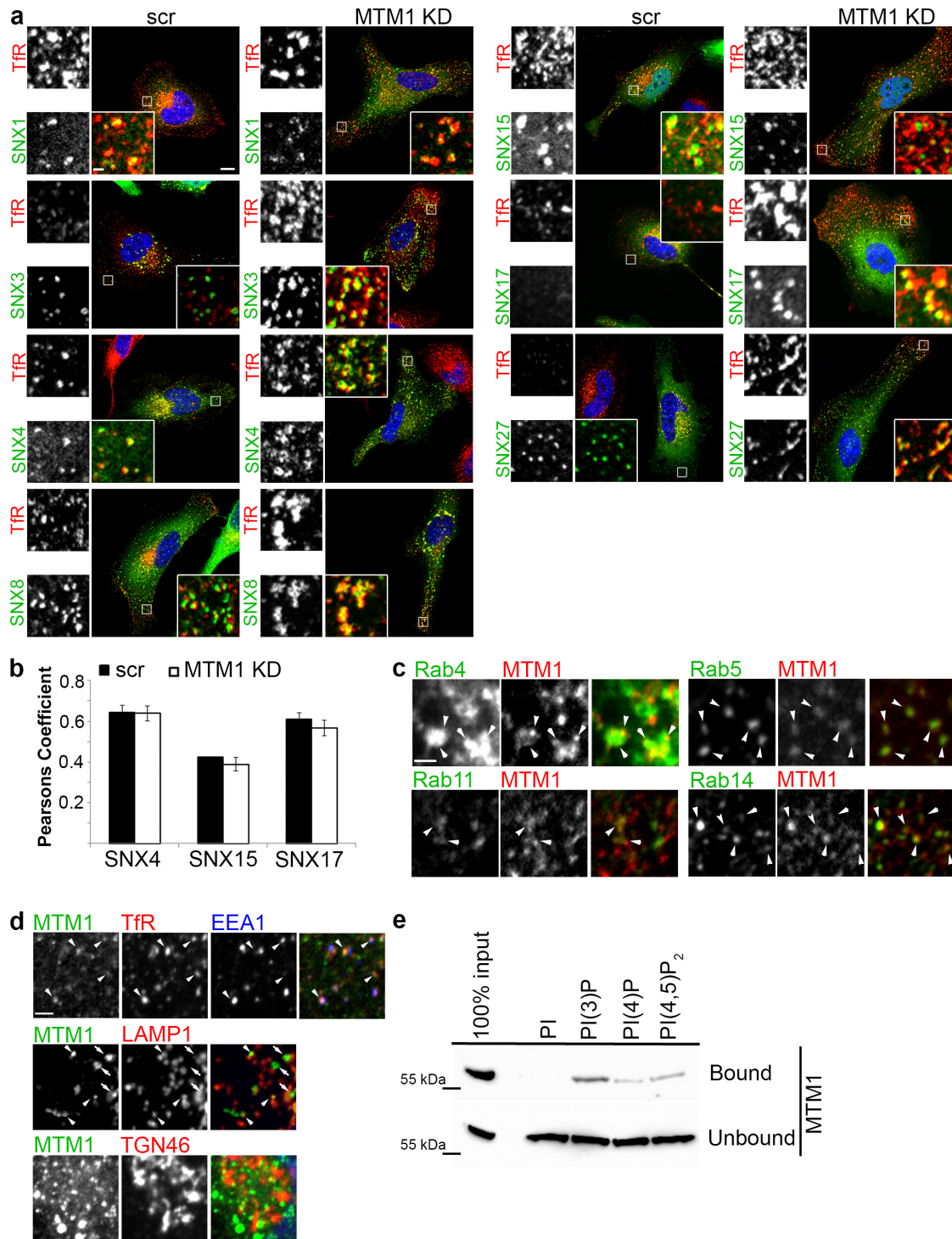
Extended Data Figure 3 | Endosomal trafficking in MTM1-depleted cells and effects of depletion of other MTM family members. **a**, EGF degradation monitored over 120 min in HeLa cells using radioactive ¹²⁵I-labelled EGF was unaltered upon loss of MTM1 (MTM1-KD) compared with scrambled (scr) treated controls (mean ± s.e.m., *n* = 3 independent experiments, each experiment done in technical triplicates). **b**, Quantitative analysis of EGFR levels normalized to cell area of (scr) control or MTM1-KD cells (EGFR per unit area; *n* = 3). **c**, Secretion of GFP-reporter construct from the Golgi complex is not affected in MTM1-depleted C1 cells. Fifteen minutes after initiating secretion, GFP is enriched in the Golgi complex. GFP levels were normalized to cell number and Golgi content (*n* = 3). **d**, LC3B level remain unchanged in HeLa cells depleted of MTM1 compared with (scr) control cells (LC3B per unit area; *n* = 3). **e**, Autophagy can be induced in HeLa cells depleted of MTM1 by starvation and BafA1 treatment, monitored by LC3B-specific antibody labelling (scale bar, 10 μm). **f**, **g**, Internalized cholera toxin (Ctx) is retrogradely transported to the Golgi complex in HeLa cells treated with scrambled (scr) siRNA or depleted of MTM1 (MTM1-KD).

f, Quantified are Pearson's coefficients between internalized Ctx-CF568 and the Golgi complex, assessed by GM130-specific antibody labelling (*n* = 3). **g**, Scale bar, 10 μm; magnified insets, 1 μm. **h**, Cholera toxin uptake quantified as the percentage of cells with internalized Ctx is unchanged upon loss of MTM1 (*n* = 3). **i**, Efficient siRNA-mediated depletion of eGFP-tagged catalytically active myotubularin-related (MTMR) proteins (MTMR1, MTMR2, MTMR4, MTMR7) from HeLa cells as demonstrated by immunoblot analysis. For blot source data, see Supplementary Fig. 1. **j**, Depletion of MTMR1, but not MTMR2, MTMR4, or MTMR7, phenocopied TfR mislocalization observed in MTM1-depleted HeLa cells (*n* = 3). **k**, Inhibition of Tf exocytosis in cells depleted of MTMR1, but not MTMR2, MTMR4, or MTMR7, compared with (scr) controls (Tf exocytic events per unit area; *n* = 3). *n*, number of independent experiments with 15–30 images (**b–d**, **f**, **h**, **j**) or five videos (**k**) analysed per condition per experiment. Mean ± s.e.m.; NS, non-significant, ***P* < 0.01, ****P* < 0.001, unpaired two-tailed *t*-test. **e**, **g**, Representative images from one of two (**e**) or three (**g**) independent experiments are shown.



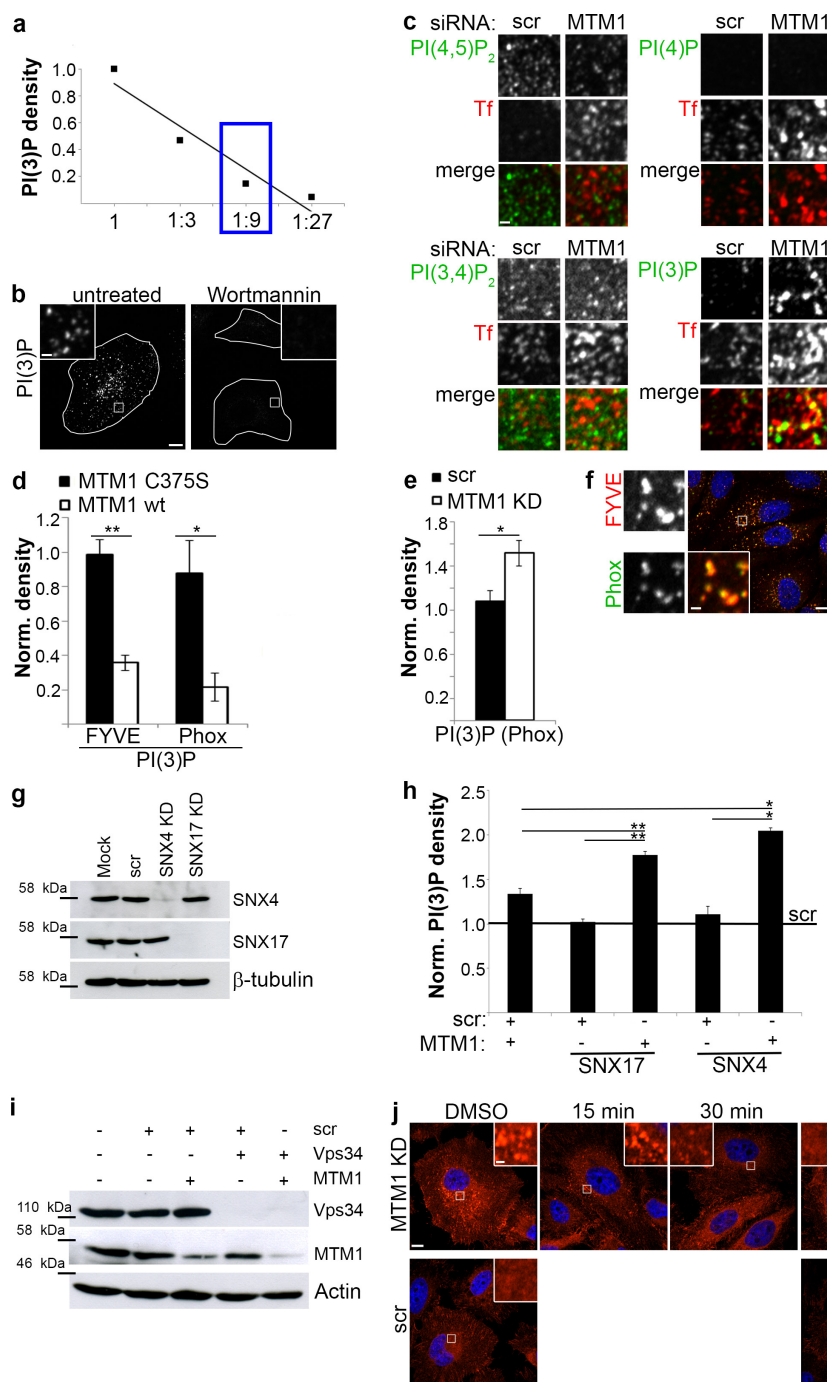
Extended Data Figure 4 | Endosomal compartments in MTM1-depleted cells. **a**, Tfr-containing endosomal tubules accumulate in HeLa cells depleted of MTM1. Scale bar, 10 µm; magnified insets, 2 µm. **b**, **c**, HeLa cells depleted of MTM1 were labelled for the indicated protein by specific antibodies. EEA1 localizes to sub-plasma-membrane Tfr accumulations, distinct from sites containing the endocytic adaptor AP-2 or the retromer component Vps26. APPL1 and LAMP1 did not localize to Tfr accumulations in MTM1-depleted cells. No changes in the GM130-containing *cis*-Golgi compartment were observed. **c**, Quantification of co-localization between the indicated protein and Tfr in the cell periphery calculated using Pearson's correlation coefficient (EEA1 and VPS26: $n = 4$; AP2 and LAMP1: $n = 3$). **d**, Distribution of Rab proteins in HeLa cells treated with scrambled (scr) siRNA or depleted of MTM1 (MTM1-KD). Confocal images of HeLa cells expressing eGFP-Rabs (Rab4, Rab5, Rab11, Rab14, Rab35). Note the strong accumulation of early and recycling endosomal Rabs on sub-plasma-membrane Tfr-containing

endosomes in MTM1-KD cells. **e**, Quantification of co-localization between the indicated Rab protein and Tfr in the cell periphery calculated using Pearson's correlation coefficient (Rab5 and Rab11: $n = 4$; Rab4 and Rab14: $n = 3$). **f**, **g**, Distinct early and recycling endosomal compartments in MTM1-depleted cells. **f**, HeLa cells treated with scrambled siRNA or depleted of MTM1 and co-expressing eGFP-Rabs (Rab4, Rab11, Rab14). Rab5 was endogenously labelled using a Rab5-specific antibody. **g**, Pearson's correlation coefficients between eGFP-Rabs and endogenous Rab5 in the cell periphery ($n = 3$). n , number of independent experiments with 15–30 images analysed per condition per experiment. Mean \pm s.e.m., unpaired two-tailed *t*-test. **a**, **b**, **d**, **f**, Representative images from 1 of 3 (**a**, **b**: for AP2, LAMP1, APPL1, GM130; **d**: for Rab4, Rab14, **f**) or 4 (**b**: for EEA1, VPS26; **d**: for Rab5, Rab11) independent experiments or from 1 experiment with 15 images per condition (**d**: Rab35) are shown. **b**, **d**, **f**, Scale bar, 10 µm; magnified insets, 1 µm.



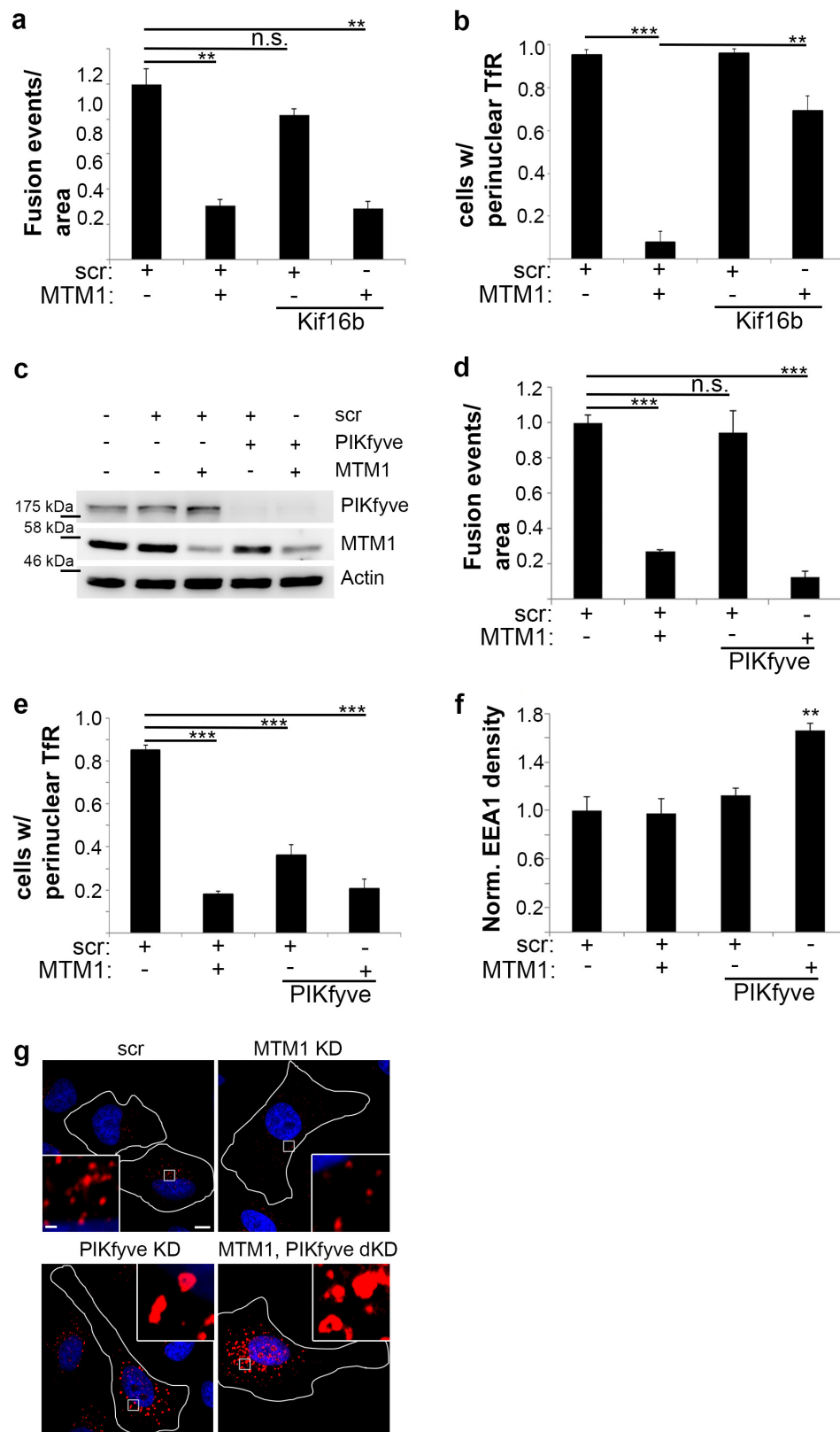
Extended Data Figure 5 | Endosomal accumulation of PI(3)P-binding sorting nexin (SNX) proteins upon loss of MTM1 and localization of MTM1 to PI(3)P-containing endosomes. **a**, Distribution of SNX proteins in HeLa cells treated with scrambled (scr) siRNA or depleted of MTM1 (MTM1-KD). Accumulation of PI(3)P-binding SNXs (eGFP-tagged SNX1, SNX3, SNX4, SNX8, SNX17, SNX27) but not eGFP-tagged SNX15 on sub-plasma-membrane TfR endosomes in MTM1-KD cells. Scale bar, 10 μ m; magnified insets, 1 μ m. **b**, Quantification of co-localization between the indicated SNX protein and TfR in the cell periphery calculated using Pearson's correlation coefficient. Mean \pm s.e.m., $n = 3$ independent experiments, unpaired two-tailed t -test. Numbers of images analysed: experiment 1 (scr: SNX4 $n = 13$, SNX15 $n = 11$, SNX17 $n = 25$; MTM1-KD: SNX4 $n = 14$, SNX15 $n = 10$, SNX17 $n = 22$); experiment 2 (scr: SNX4 $n = 15$, SNX15 $n = 14$, SNX17 $n = 15$; MTM1-KD: SNX4 $n = 15$, SNX15 $n = 15$, SNX17 $n = 14$); experiment 3 (scr: SNX4 $n = 15$, SNX15 $n = 15$, SNX17 $n = 15$; MTM1-KD: SNX4 $n = 9$, SNX15 $n = 20$,

SNX17 $n = 20$). **c**, HeLa cells co-expressing eGFP-Rabs and mCherry-MTM1. Co-localization of MTM1 with early and recycling endosomal Rabs (Rab4, Rab5, Rab11, Rab14) is indicated by arrowheads. Scale bar, 2 μ m. **d**, HeLa cells expressing eGFP-MTM1 were labelled for the indicated endosomal markers by specific antibodies. Arrowheads mark co-localization of MTM1 with the indicated proteins, whereas arrows mark MTM1-positive endosomes devoid of the indicated protein. Scale bar, 2 μ m. **e**, Binding of HA-tagged MTM1 expressed in Hek293 cells to liposomes containing 5 mol% of the indicated phosphoinositide in flotation assays. MTM1 predominantly binds to PI(3)P-containing liposomes. Input, 30% (top) for bound and 15% (bottom) for unbound fractions. For blot source data, see Supplementary Fig. 1. **a**, **c**, **d**, Representative images from 1 of 3 (**a**, for SNX4, SNX15, SNX17) independent experiments or from 1 experiment with 6–20 images per condition (**a**, for SNX3, SNX8, SNX27, **c**, **d**) are shown.



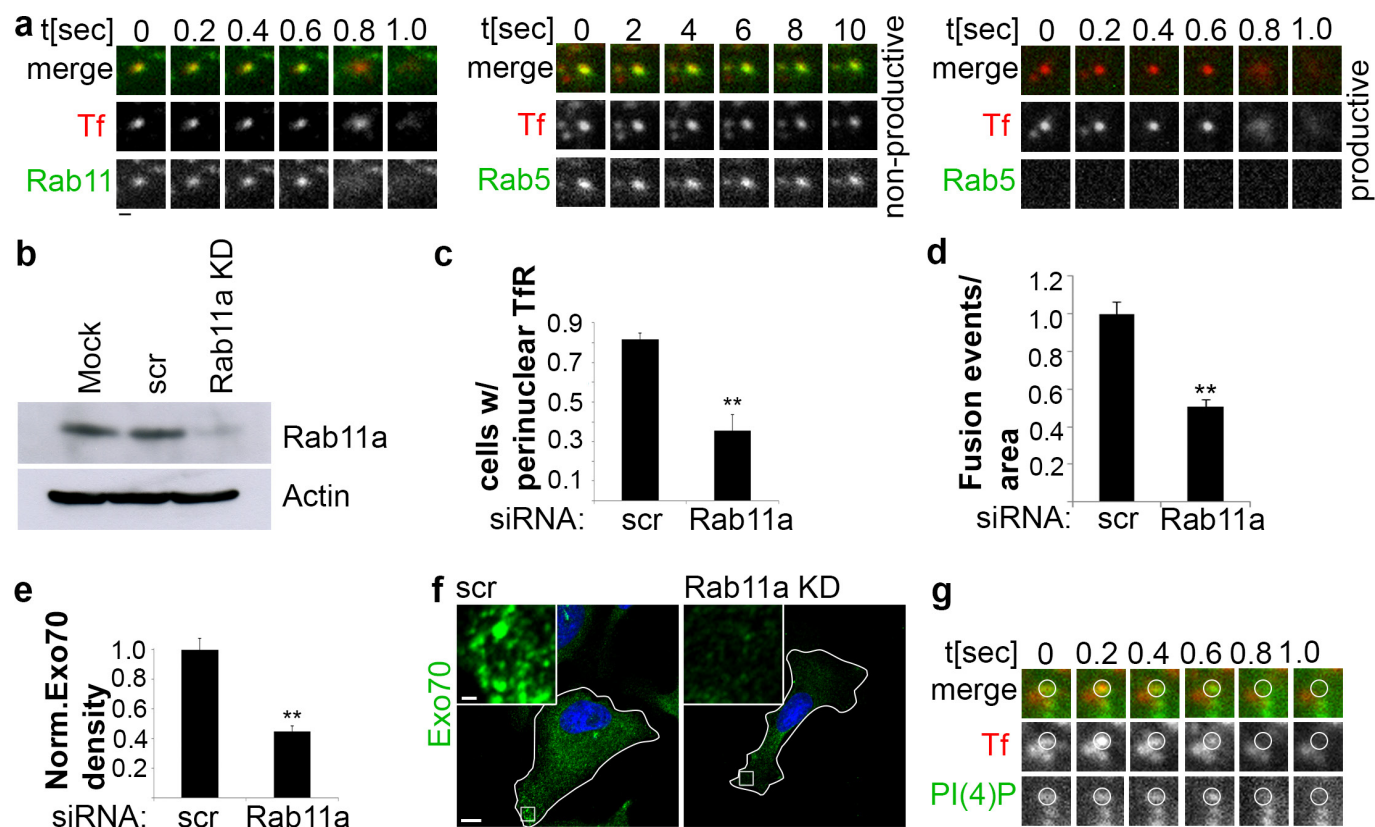
Extended Data Figure 6 | Sub-plasma-membrane endosomal TfR accumulations are selectively enriched in PI(3)P. **a**, HeLa cells were labelled for PI(3)P using different eGFP-2xFYVE domain concentrations, starting with $2.25 \mu\text{g ml}^{-1}$ (1) and subsequent 1:3 dilutions of (1) ($0.75 \mu\text{g ml}^{-1}$ (1:3), $0.25 \mu\text{g ml}^{-1}$ (1:9), $0.083 \mu\text{g ml}^{-1}$ (1:27)). PI(3)P levels per cell area were quantified. The titration curve is shown and the linear range indicated by the blue box. **b**, HeLa cells treated with DMSO or $2 \mu\text{M}$ wortmannin were stained for PI(3)P using purified eGFP-2xFYVE. Note the absence of eGFP-2xFYVE staining in cells treated with the broad-spectrum phosphoinositide 3-kinase inhibitor wortmannin. **c**, PIPx stainings in MTM1-depleted HeLa cells: eGFP-2xFYVE was used to detect PI(3)P. PI(4)P, PI(3,4)P₂ and PI(4,5)P₂ were detected with specific antibodies against the corresponding PIPx species. Sub-plasma-membrane TfR accumulations are selectively enriched in PI(3)P. **d**, Decreased PI(3)P levels in HeLa cells overexpressing WT but not mutant inactive (C375S) mCherry-MTM1. Depletion of PI(3)P is consistent with PI(3)P 3-phosphatase activity of MTM1. Cells were labelled for PI(3)P using eGFP-2xFYVE domain or Phox-AF488 as indicated and the relative levels of PI(3)P normalized to cell area were quantified ($n = 3$). **e**, Increase of

PI(3)P in MTM1-depleted HeLa cells. Cells were labelled for PI(3)P using Phox-AF488 and the relative levels of PI(3)P normalized to cell area were quantified ($n = 3$). **f**, HeLa cells were labelled for PI(3)P using eGFP-2xFYVE domain and Phox-AF488 as indicated. **g**, Efficient silencing of SNX4 and SNX17 in HeLa cells as demonstrated by immunoblotting. **h**, Co-depletion of MTM1 and SNX4 or SNX17 reveals additional pools of PI(3)P inaccessible in cells only depleted of MTM1. Relative levels of PI(3)P in cells treated with the indicated siRNAs ($n = 3$). **i**, Efficient co-depletion of Vps34 and MTM1 in HeLa cells as demonstrated by immunoblotting. **j**, MTM1-KD cells or (scr) controls were treated with VPS34-IN1 for the indicated times and labelled for β 1-integrin. Pharmacological inhibition of Vps34 rescues β 1-integrin accumulation in MTM1-depleted HeLa cells within 60 min, assessed by β 1-integrin-specific antibody labelling. n , number of independent experiments with 15–30 images analysed per condition per experiment. Mean \pm s.e.m., * $P < 0.05$, ** $P < 0.01$, unpaired two-tailed t -test. **b**, **c**, **f**, **j**, Representative images from 1 of 3 (**c**) independent experiments or from 1 experiment with 10–15 images per condition (**b**, **f**, **j**) are shown. **b**, **c**, **f**, **j**, Scale bar, $10 \mu\text{m}$; magnified insets, $1 \mu\text{m}$. For blot source data, see Supplementary Fig. 1.



Extended Data Figure 7 | PI(3)P manipulations. **a**, Co-depletion of MTM1 and Kif16b does not restore defective Tf exocytosis in MTM1-depleted cells (Tf exocytic events per unit area; $n = 3$). **b**, Co-depletion of MTM1 and Kif16b restores perinuclear TfR localization. Normalized fraction of HeLa cells with perinuclear TfR in cells treated with the indicated siRNAs ($n = 3$). **c**, Efficient co-depletion of PIKfyve and MTM1 in HeLa cells as demonstrated by immunoblotting. For blot source data, see Supplementary Fig. 1. **d**, Co-depletion of MTM1 and PIKfyve does not restore defective Tf exocytosis in MTM1-depleted cells (Tf exocytic events per unit area; $n = 3$). **e**, Co-depletion of PIKfyve and MTM1 is

unable to restore perinuclear TfR localization. Normalized fraction of cells with perinuclear TfR in cells treated with the indicated siRNAs ($n = 3$). **f**, Co-depletion of PIKfyve and MTM1 leads to increased EEA1 density ($n = 3$). **g**, Swelling of EEA1-positive endosomes in PIKfyve-depleted and MTM1/PIKfyve co-depleted cells. Scale bar, 10 μm ; magnified insets, 1 μm . Representative images from one of three independent experiments are shown. n , number of independent experiments with 15–30 images (**b**, **e**, **f**) or 5 videos (**a**, **d**) analysed per condition per experiment. Mean \pm s.e.m.; NS, non-significant, $**P < 0.01$, $***P < 0.001$, unpaired two-tailed t -test.

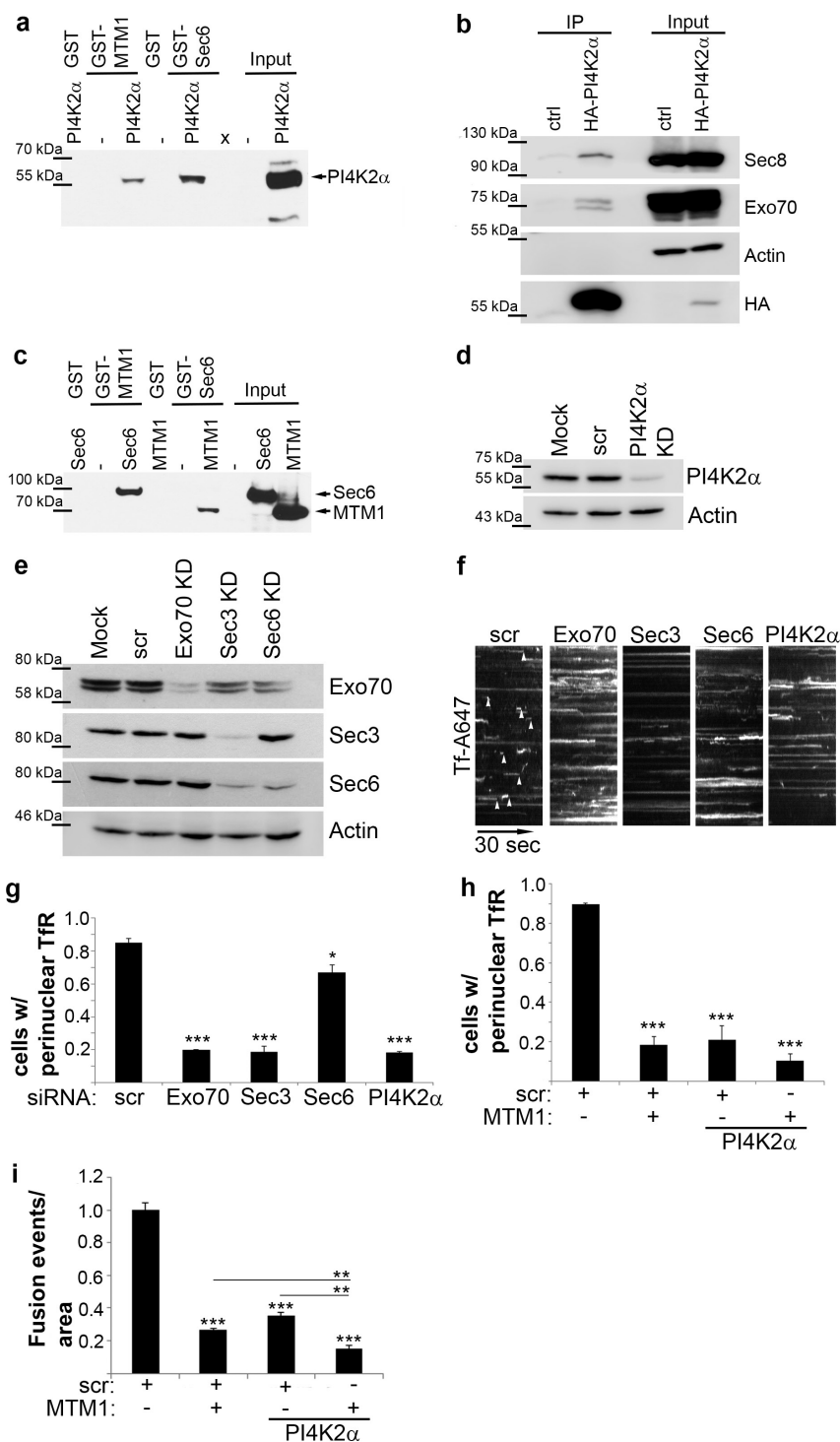


Extended Data Figure 8 | Rab11a in endosomal exocytosis of Tf.

a, Tf exocytosis was monitored by dual-colour TIRF microscopy in HeLa cells expressing GFP-Rab5a or GFP-Rab11a. Snapshots of representative Tf-positive, Rab-containing endosomes are shown. Scale bar, 400 nm. **b**, Efficient siRNA-mediated depletion of Rab11a from HeLa cells as demonstrated by immunoblot analysis. **c**, TfR mislocalizes upon loss of Rab11a in HeLa cells. Quantified was normalized fraction of cells with perinuclear TfR ($n = 3$). **d**, Impaired Tf exocytosis monitored by TIRF microscopy in Rab11a-depleted HeLa cells (Tf exocytic events per unit area; $n = 3$). **e**, Reduced endosomal exocyst association in Rab11a-depleted HeLa cells revealed by confocal imaging using Exo70-specific antibody labelling ($n = 3$). **f**, Scale bar, 10 μm ; magnified insets, 1 μm . Representative

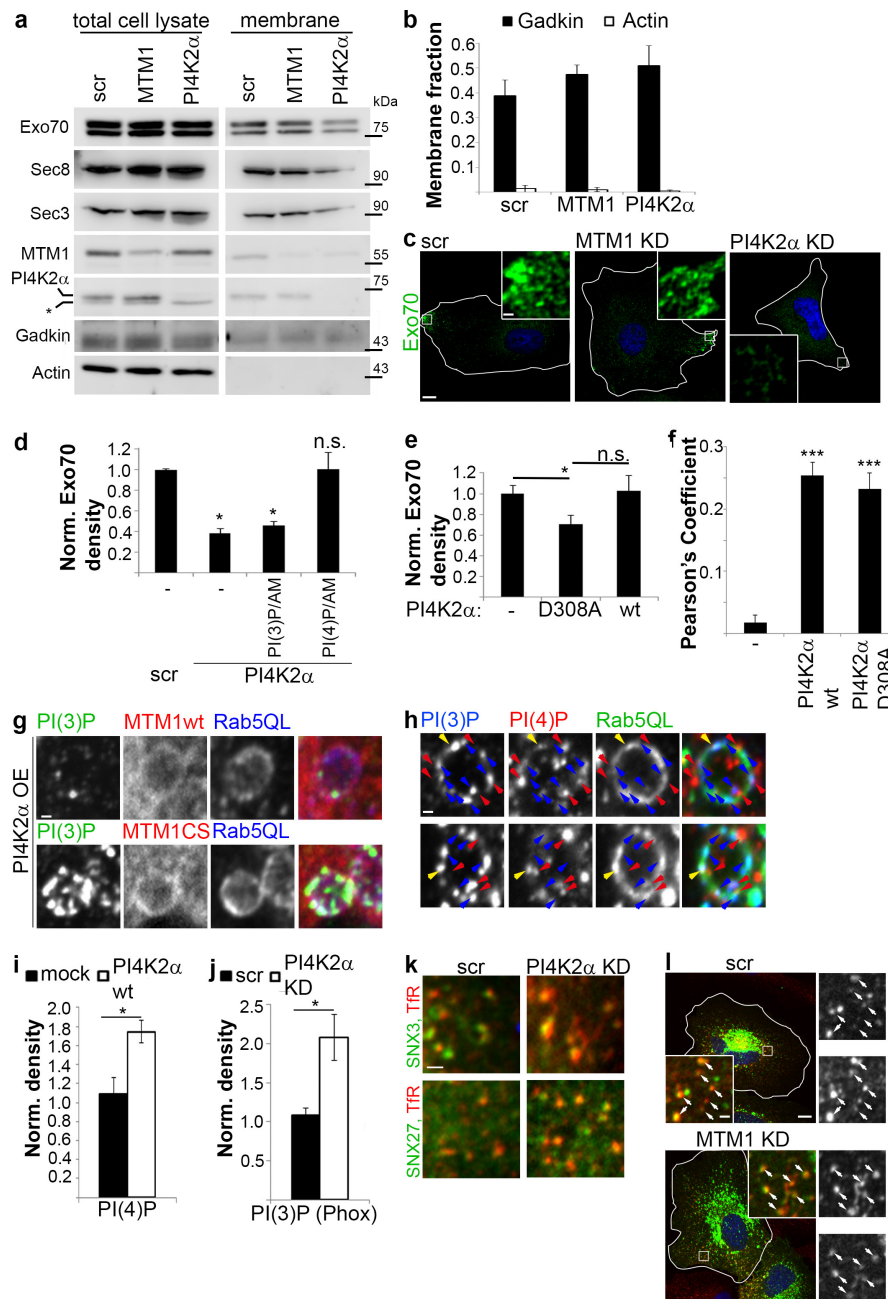
images from one of three independent experiments are shown.

g, Tf-Alexa647 (red) exocytosis from endosomes in HeLa cells overexpressing the PI(4)P reporter eGFP-2xPH-FAPP1 (green) analysed by dual-colour TIRF microscopy. Order of acquisition was reversed compared with Fig. 4a to exclude imaging artefacts. The 647 nm channel was imaged before acquisition of the 488 nm channel. Scale bar, 400 nm. n , number of independent experiments with 15–30 images (**c**, **e**) or 5 videos (**d**) analysed per condition per experiment. Mean \pm s.e.m., $***P < 0.01$, unpaired two-tailed t -test. **a**, **g**, Representative images of one experiment with ten videos per condition (**a**) or two independent experiments with ten videos per condition per experiment (**g**) are shown.



Extended Data Figure 9 | PI4K2α and exocyst associate and are required for endosomal exocytosis of Tf. **a**, PI4K2α interacts directly with Sec6 and MTM1. Pull down assays were performed using GST-MTM1 or GST-Sec6 as bait to pull down HA-tagged PI4K2α from Cos-1 cells and detected using PI4K2α-specific antibody. The PI4K2α-specific band is indicated by the arrow. **b**, Co-immunoprecipitation of PI4K2α and exocyst components from stable Hek293. **c**, Reciprocal interaction between MTM1 and Sec6. Pull down assays were performed using GST-MTM1 as bait to pull down B10-tagged Sec6 or using GST-Sec6 to pull down B10-tagged MTM1 from Cos-1 cells and detected using B10-specific antibodies as indicated by arrows. **d**, **e**, Efficient depletion of PI4K2α (**d**) and exocyst subunits Exo70, Sec3, and Sec6 (**e**) as demonstrated by immunoblotting of HeLa cell lysates. **f**, Depletion of exocyst or PI4K2α from HeLa cells impairs Tf exocytosis monitored by TIRF microscopy. Kymographs of

Tf fluorescence signal over 28 μm; arrowheads indicate time of fusion. Representative images from one of three independent experiments are shown. **g**, Depletion of PI4K2α or exocyst from HeLa cells causes TfR mislocalization. Normalized fraction of cells with perinuclear TfR was quantified ($n = 4$). **h**, Co-depletion of MTM1 and PI4K2α does not restore perinuclear TfR localization in MTM1-depleted HeLa cells. Normalized fraction of cells with perinuclear TfR in cells treated with the indicated siRNAs ($n = 3$). **i**, Co-depletion of MTM1 and PI4K2α does not restore defective Tf exocytosis in MTM1-depleted HeLa cells (Tf exocytic events per unit area; $n = 3$). n , number of independent experiments with 15–30 images (**g**, **h**) or 5 videos (**i**) analysed per condition per experiment. Mean \pm s.e.m., * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, unpaired two-tailed t -test. For blot source data, see Supplementary Fig. 1.



Extended Data Figure 10 | Exocyst and MTM1 recruitment to endosomes depends on PI4K2α and initiates MTM1-dependent PI(3)P-dephosphorylation. **a**, Depletion of PI4K2α but not of MTM1 impairs membrane recruitment of Exo70, Sec8, Sec3, and MTM1. Representative immunoblot is shown. Asterisk indicates non-specific band recognized by PI4K2α antibodies. **b**, Efficient membrane-cytosol fractionation verified by quantifying the membrane/total protein ratio of the palmitoylated membrane protein gadkin and cytosolic actin (mean ± s.e.m., $n = 5$ independent experiments). **c**, Depletion of PI4K2α but not of MTM1 in HeLa cells leads to decreased Exo70 level. Scale bar, 10 μm; magnified insets, 1 μm. **d**, Reduced endosomal exocyst association in PI4K2α-depleted cells is rescued by membrane-permeant PI(4)P but not PI(3)P revealed by confocal imaging using Exo70-specific antibody labelling ($n = 4$). **e**, Expression of mutant inactive (D308A) but not WT PI4K2α in HeLa cells reduces endosomal exocyst association revealed by confocal imaging using Exo70-specific antibody labelling ($n = 3$). **f**, PI4K2α-dependent recruitment of MTM1 is independent of its kinase activity. Expression of WT or inactive mutant (D308A) eGFP-PI4K2α recruits mCherry-MTM1 to Rab5 Q79L endosomes. Quantified are Pearson's coefficients between mCherry-MTM1 and HA-Rab5 Q79L ($n = 4$). **g**, HeLa cells co-expressing WT or inactive, mutant (C375S) MTM1, PI4K2α, and Rab5 Q79L were labelled for PI(3)P using

eGFP-2xFYVE. Inactive MTM1 localizes to PI(3)P-positive early endosomes, while this PI(3)P pool is lost upon recruitment of active MTM1. Scale bar, 1 μm. **h**, PI(3)P and PI(4)P localize to distinct subdomains on Rab5 Q79L endosomes labelled using eGFP-2xFYVE and a PI(4)P-specific antibody respectively. PIP localizations are indicated by arrowheads (blue: PI(3)P; red: PI(4)P; yellow: co-localization of PI(3)P and PI(4)P). Scale bar, 1 μm. **i**, Increased PI(4)P levels in HeLa cells overexpressing eGFP-PI4K2α WT compared with mock transfected cells. Increase of PI(4)P is consistent with phosphoinositide 4-kinase activity of PI4K2α. Shown are the relative levels of PI(4)P normalized to cell area ($n = 3$). **j**, Increase of PI(3)P in PI4K2α-depleted HeLa cells. Cells were labelled for PI(3)P using Phox-AF488 and the relative levels of PI(3)P normalized to cell area quantified ($n = 3$). **k**, Recruitment of eGFP-SNX3 and eGFP-SNX27 to dispersed endosomes in PI4K2α-depleted HeLa cells. Scale bar, 2 μm. **l**, eGFP-PI4K2α expressed in MTM1-depleted HeLa cells localizes to TfR-positive endosomes, assessed by TfR-specific antibody labelling and indicated by arrows. Scale bar, 10 μm; magnified insets, 1 μm. **n**, number of independent experiments with 15–30 images (**d**, **e**, **f**, **i**, **j**) analysed per condition per experiment. Mean ± s.e.m.; NS, non-significant, $*P < 0.05$, $***P < 0.001$, unpaired two-tailed t -test. **c**, **g**, **h**, **k**, **l**, Representative images from 1 of 4 (**c**, **l**) independent experiments or from 1 experiment with 10–20 images per condition (**g**, **h**, **k**) are shown.

Response and resistance to BET bromodomain inhibitors in triple-negative breast cancer

Shaokun Shu^{1,2*}, Charles Y. Lin^{1,2*}, Housheng Hansen He^{1,2,3,4,5*}, Robert M. Witwicki^{1,2*}, Doris P. Tabassum¹, Justin M. Roberts¹, Michalina Janiszewska^{1,2}, Sung Jin Huh^{1,2}, Yi Liang⁴, Jeremy Ryan^{1,2}, Ernest Doherty^{1,6}, Hisham Mohammed⁷, Hao Guo³, Daniel G. Stover^{1,2}, Muhammad B. Ekram^{1,2}, Guillermo Peluffo^{1,2}, Jonathan Brown^{1,2}, Clive D'Santos⁷, Ian E. Krop^{1,2}, Deborah Dillon^{1,8}, Michael McKeown^{1,2}, Christopher Ott^{1,2}, Jun Qi^{1,2}, Min Ni^{1,2}, Prakash K. Rao⁹, Melissa Duarte⁹, Shwu-Yuan Wu¹⁰, Cheng-Ming Chiang¹⁰, Lars Anders¹¹, Richard A. Young¹¹, Eric P. Winer^{1,2}, Antony Letai^{1,2}, William T. Barry^{2,3}, Jason S. Carroll⁷, Henry W. Long^{1,9}, Myles Brown^{1,2,9}, X. Shirley Liu^{3,9,12}, Clifford A. Meyer³, James E. Bradner^{1,2,12} & Kornelia Polyak^{1,2,9,12}

Triple-negative breast cancer (TNBC) is a heterogeneous and clinically aggressive disease for which there is no targeted therapy^{1–3}. BET bromodomain inhibitors, which have shown efficacy in several models of cancer^{4–6}, have not been evaluated in TNBC. These inhibitors displace BET bromodomain proteins such as BRD4 from chromatin by competing with their acetyl-lysine recognition modules, leading to inhibition of oncogenic transcriptional programs^{7–9}. Here we report the preferential sensitivity of TNBCs to BET bromodomain inhibition *in vitro* and *in vivo*, establishing a rationale for clinical investigation and further motivation to understand mechanisms of resistance. In paired cell lines selected for acquired resistance to BET inhibition from previously sensitive TNBCs, we failed to identify gatekeeper mutations, new driver events or drug pump activation. BET-resistant TNBC cells remain dependent on wild-type BRD4, which supports transcription and cell proliferation in a bromodomain-independent manner. Proteomic studies of resistant TNBC identify strong association with MED1 and hyper-phosphorylation of BRD4 attributable to decreased activity of PP2A, identified here as a principal BRD4 serine phosphatase. Together, these studies provide a rationale for BET inhibition in TNBC and present mechanism-based combination strategies to anticipate clinical drug resistance.

To explore non-oncogene addiction to BRD4 in breast cancer, we studied a series of BET bromodomain inhibitors (BBI) across breast cell lines reflecting transcriptionally defined breast cancer subtypes: luminal, HER2⁺ and TNBC^{2,10}, as well as MCF10A and MCF12A basal/mesenchymal immortalized mammary epithelial cells (Supplementary Table 1). Potent inhibitory effects were observed preferentially in TNBC lines, compared to more resistant luminal lines (Fig. 1a). Analysis of potency of drug response and subtype or known driver mutations identified the basal subtype as the only significant association ($P = 0.0475$) (Supplementary Table 1 and data not shown). BRD4 dependency was confirmed by RNA interference and phenocopied BBI (Extended Data Fig. 1a–c). JQ1 or BRD4 knockdown induced growth inhibition and resulted in G1 arrest and apoptosis (Extended Data Fig. 1d–g). Expression of factors described to mediate JQ1 effect (MYC) or required for TNBC growth (JAK2/STAT3) showed no clear association with JQ1 sensitivity (Extended Data Fig. 1h and Extended Data Fig. 2a, b). JQ1 treatment of TNBC cells induced significant morphologic

changes consistent with induction of senescence, confirmed by β -galactosidase staining and luminal differentiation evidenced by changes in the expression of basal and luminal markers (Extended Data Fig. 2c, d and Fig. 1b).

Extending the translational significance of these findings, we evaluated the ability of JQ1 to inhibit tumour growth in murine TNBC xenografts. Two-week treatment efficiently inhibited established tumour growth from SUM159 and MDA-MB-231 lines, and patient-derived primary human TNBC xenografts (Fig. 1c and Extended Data Fig. 2e, f). Downregulation of BRD4 using two independent TET-inducible short hairpin RNAs produced even more pronounced effects, leading to complete tumour regression and failure to regrow even after discontinuing doxycycline treatment (Fig. 1c and Extended Data Fig. 2g). Evidence of BBI-induced basal-to-luminal differentiation was confirmed *in vivo* (Extended Data Fig. 2f, h).

Using integrated epigenomic analysis (Supplementary Table 2), we identified the direct transcriptional targets of BBI in TNBC. BBI binding was identified at active promoter and enhancer regions using ChEM-seq¹¹ for biotinylated JQ1 (Bio-JQ1) enrichment and chromatin immunoprecipitation followed by sequencing (ChIP-seq) for acetyl-histone (H3K27ac) and BRD4 enrichment, with the three marks showing near perfect co-localization (Fig. 1d and Extended Data Fig. 3a). BBI efficiently displaced chromatin-bound BRD4 in treated SUM159 (Fig. 1e and Extended Data Fig. 3b) and in SUM149 cells (Extended Data Fig. 3c). To identify biologically relevant, direct targets of BBI in SUM159 and SUM149 cells, we quantified binding of Bio-JQ1 and BRD4 genome-wide and found strong enrichment at 219 and 159 super-enhancers, respectively (Fig. 1f, Extended Data Fig. 3d and Supplementary Table 3)^{8,9,12,13}. Transcription factors with known roles in breast cancer, such as POU5F1B/MYC¹⁴ and HIF1 α ¹⁵, were evident among top super-enhancer-associated genes in both lines. Kinetic effects of JQ1 treatment on gene expression demonstrated preferential super-enhancer-associated gene downregulation (Fig. 1g and Extended Data Fig. 3e, f). Expression changes were observed within 3 h after JQ1 treatment and, as expected, more genes were significantly down- than upregulated (Extended Data Fig. 3g–j and Supplementary Table 4). Unsupervised Metacore¹⁶ analysis of JQ1-affected target pathways revealed downregulation of regulatory and effector genes in anti-apoptotic and JAK/STAT signalling pathways (Extended Data Fig. 3k).

¹Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. ²Department of Medicine, Brigham and Women's Hospital, and Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. ³Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, and Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, USA. ⁴Princess Margaret Cancer Center/University Health Network, Toronto, Ontario M5G1L7, Canada. ⁵Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G2M9, Canada. ⁶Harvard University, Cambridge, Massachusetts 02138, USA. ⁷Cancer Research UK, Cambridge Institute, University of Cambridge, Cambridge CB2 0RE, UK. ⁸Department of Pathology, Brigham and Women's Hospital, and Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁹Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. ¹⁰Simmons Comprehensive Cancer Center, Departments of Biochemistry and Pharmacology, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ¹¹Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA. ¹²Broad Institute, Cambridge, Massachusetts 02142, USA.

*These authors contributed equally to this work.

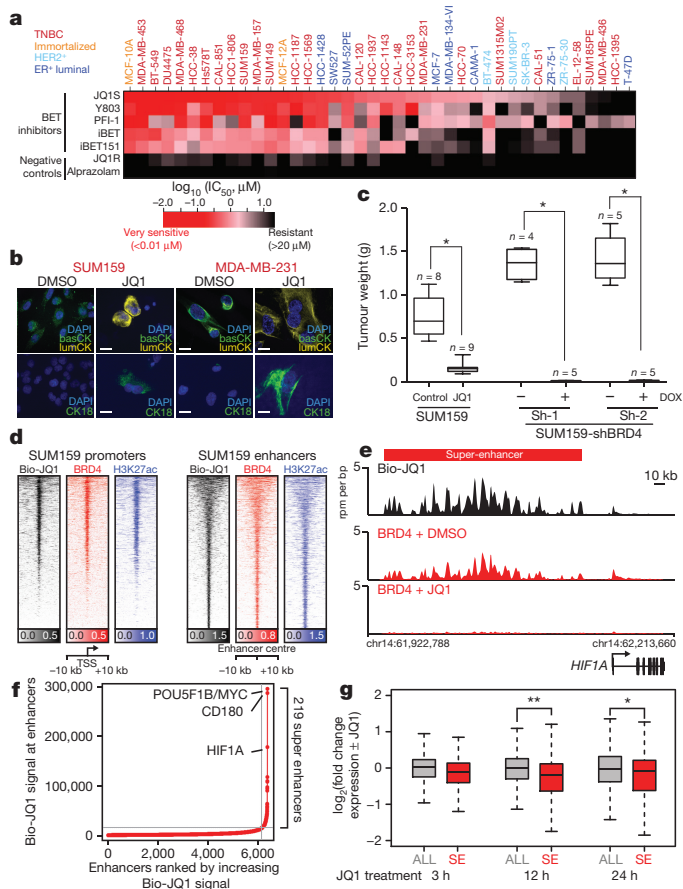


Figure 1 | Response to BBIs in breast cancer. **a**, Heat map of mean IC_{50} s of BBIs and inactive analogues in breast cell lines. Error bars represent s.e.m. **b**, Immunofluorescence of basal (basCK) and luminal (lumCK and CK18) cytokeratins in TNBC lines. Scale bars, 20 μ m. **c**, Box plots depicting xenograft weights; n indicates the number of mice per experiment. $*P < 0.0001$ (unpaired t -test). **d**, Heat map showing biotinylated JQ1 (Bio-JQ1), BRD4 and H3K27ac binding at transcription start site (TSS) and Bio-JQ1-bound enhancer regions. Each row represents a single genomic region (± 10 kb) from TSS or enhancer centre. Genomic occupancy is shaded by binding intensity in units of reads per million per base pair (rpm per bp). **e**, Gene tracks depicting Bio-JQ1 and BRD4 with or without JQ1 in SUM159 cells at the *HIF1A* locus. x-axis, chromosome position with gene structures below; y-axis, genomic occupancy in units of rpm per bp; red bar, *HIF1A* super-enhancer. **f**, Plot of enhancers defined in untreated SUM159 cells ranked by increasing Bio-JQ1 signal (units rpm). Grey line marks cut-off discriminating typical from super-enhancers. **g**, Box plots showing the \log_2 fold change in expression relative to control of either all active or super-enhancer (SE)-associated genes upon JQ1 treatment. $**P < 10^{-5}$ (Welsh's t -test); $*P < 10^{-3}$ (Welsh's t -test).

These data support selective disruption of super-enhancer-associated genes by JQ1, leading to deregulation of coordinated transcriptional pathways involved in cell proliferation, invasion and survival.

Dissecting resistance to targeted therapy is critical to elucidate mechanisms of drug and target action, and to suggest approaches to treat or anticipate drug resistance in patients. Therefore, we established BBI-resistant TNBC cell lines by long-term culture of both SUM159 and SUM149 cells in escalating JQ1 doses. Low (0.5 μ M) and high (2.0 μ M) doses of JQ1 severely impaired proliferation of parental SUM159 and SUM149 lines, reducing viable cells after 6 days (Fig. 2a and Extended Data Fig. 3l). In contrast, JQ1-resistant cells (SUM159R and SUM149R) proliferated linearly, even in high JQ1 doses (20 μ M) (Fig. 2a and Extended Data Fig. 3l). BBI-resistance is not attributable to drug export, as MDR1 and other transporters are not transcriptionally upregulated (Extended Data Fig. 4a), co-incubation with MDR1

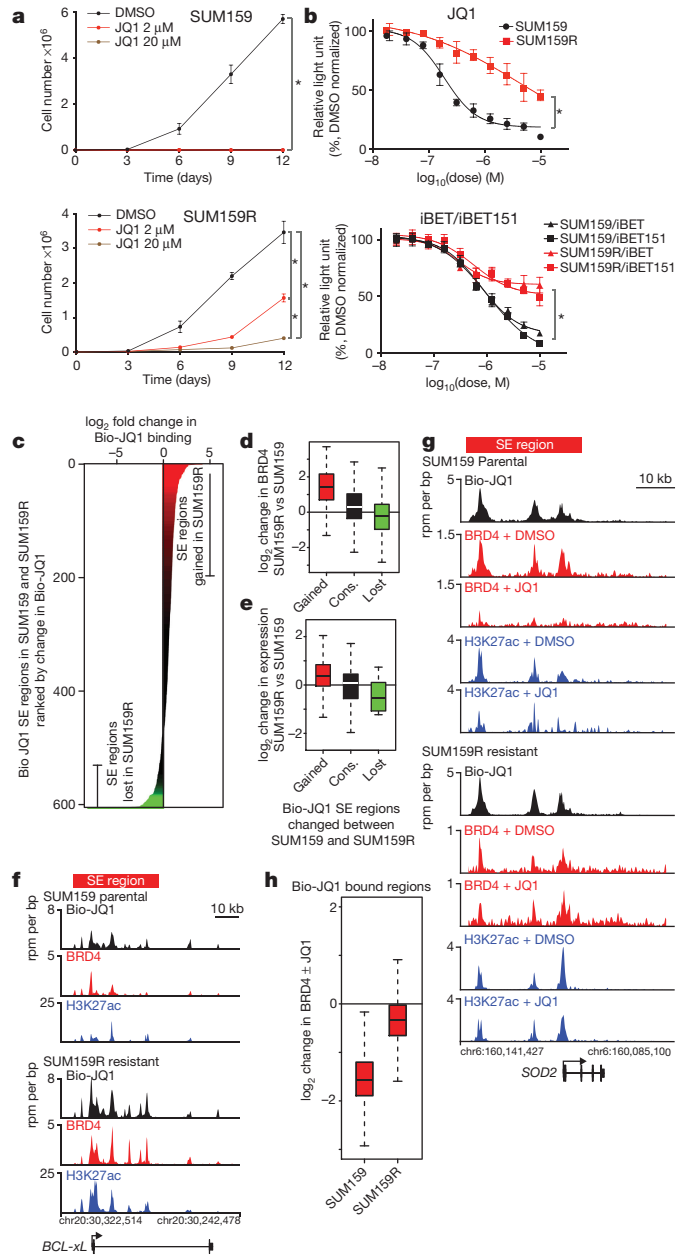


Figure 2 | Acquired BBI-resistance in TNBC. All error bars represent s.d., $n = 3$. **a**, Viable cell numbers after JQ1 treatment. $*P < 0.0001$ (two-way ANOVA). **b**, Cellular viability after treatment with BBIs. $*P < 0.0001$ (non-linear regression, extra sum-of-squares test). **c**, Genomic regions containing a super-enhancer in SUM159 or SUM159R cells ranked by \log_2 change in Bio-JQ1 genomic binding signal. x-axis, \log_2 fold change in Bio-JQ1 signal coloured by intensity of change. **d**, **e**, Box plot showing the \log_2 fold change in BRD4 genomic occupancy (**d**) and gene expression (**e**) at regions with gained, conserved, or lost Bio-JQ1 binding in SUM159R versus SUM159 cells. **f**, **g**, Gene tracks depicting Bio-JQ1, BRD4, and H3K27ac at the *BCL-xL* (also known as *BCL2L1*; **f**) and *SOD2* (**g**) locus. The x-axis shows position along the chromosome with gene structures drawn below. The y-axis shows genomic occupancy in units of rpm per bp. **h**, Box plot showing the \log_2 fold change in BRD4 genomic occupancy at regions bound by Bio-JQ1.

inhibitors (verapamil) had no effect (Extended Data Fig. 4b), and structurally divergent BBIs are equally inactive as JQ1 (Fig. 2b). Further support is provided by the equivalent chromatin engagement of BRD4 in sensitive and resistant cells, demonstrated by binding with Bio-JQ1 (Extended Data Fig. 4c). Notably, BBI-resistant TNBC cells retain sensitivity to compounds from orthogonal active drug classes, such as

CXCR2 and JAK2 inhibitors¹⁷, establishing specific resistance to BBIs (Extended Data Fig. 4d). Adaptive drug resistance was not attributable to outgrowth of a minor subpopulation of pre-existing resistant cells, as 10 independent single-cell-derived clones showed similar resistance profiles to pooled SUM159R cells (Extended Data Fig. 4e). Similar results were obtained *in vivo*, as SUM159R-derived xenografts were JQ1-unresponsive (Extended Data Fig. 4f). In all resistant TNBC populations studied, exome sequencing failed to identify alterations in BET bromodomain-encoding genes (for example, gatekeepers) or known driver genes (parallel pathway activation) (Supplementary Table 5).

Absent new genetic alterations, we explored the plausibility of an epigenomic mechanism of resistance. Differential enhancer analysis revealed a significant gain of super-enhancers in resistant SUM159R cells (Fig. 2c and Supplementary Table 6). The gain of Bio-JQ1 super enhancers was associated with enrichment for BRD4 binding to these genomic loci (Fig. 2d) and increased transcription of associated genes (Fig. 2e). An upstream/intragenic region of H3K27ac at the BCL-xL locus featured prominently among top gained super enhancers in SUM159R (Fig. 2f), consistent with increased BCL-xL messenger RNA and protein expression in resistant cells (Supplementary Table 7, Extended Data Fig. 4g). Functionally, cells with acquired resistance to BBI featured a concordant switch in JQ1 anti-apoptotic response based on dynamic BH3 profiling^{18,19} (Extended Data Fig. 4h).

Observing emergent enhancers in resistant cells, we assessed whether BBI-resistant TNBC cells retained non-oncogene addiction to BRD4. Notably, we observed loss of SUM159R cell viability upon BRD4 knockdown (Extended Data Fig. 5a, b). Together these studies establish persistence of BRD4 addiction despite resistance to bromodomain inhibition, establishing the plausibility of bromodomain-independent recruitment of BRD4 to enhancers in BBI-resistant TNBCs. To test this hypothesis, we performed BRD4 ChIP-seq on sensitive and resistant cells with and without JQ1. JQ1 neither displaced BRD4 from chromatin in SUM159R (Fig. 2g), nor meaningfully influenced epigenome structure by H3K27ac ChIP-seq (Extended Data Fig. 5c–g). Notably, several luminal markers (FOXA1, CD24, and luminal cytokeratins) were elevated in SUM159R cells in cell culture and *in vivo* (Extended Data Fig. 5h, i), supporting a model whereby resistance arises via essential BRD4 recruitment to chromatin in a bromodomain-independent manner. Similar observations were made in SUM149R cells and in TNBC cells inherently resistant to JQ1 (Extended Data Fig. 3h–j; Extended Data Fig. 6a–d), suggesting a general mechanism of epigenomic resistance to BBI.

To disclose potential differences in BRD4-associated complexes between sensitive and resistant SUM159 cells, we performed quantitative proteomics using RIME (rapid immunoprecipitation mass spectrometry of endogenous proteins)²⁰ with and without JQ1. Analysis of BRD4-associated proteins identified relative enrichment of MED1 and BRD3 in JQ1-treated resistant cells (Fig. 3a, Extended Data Fig. 7 and Supplementary Table 8). BRD4 immunoprecipitation followed by immunoblot for MED1 and BRD3 revealed that JQ1 efficiently displaced BRD4 from MED1 in sensitive cells, but not in resistant cells (Fig. 3b), a result confirmed in SUM149R and BBI-resistant SUM149R, as well as inherently resistant TNBC and luminal lines (Extended Data Fig. 8a). Although elevated BRD3 abundance was observed in SUM159R cells, increased association of BRD4 and BRD3 was not confirmed by immunoblot (Fig. 3b). To assess functionally whether increased recruitment of BRD4 to chromatin by MED1 underlies resistance to JQ1, we expressed an exogenous bromodomain-inactivated mutant (BDmut) with concomitant knockdown of endogenous BRD4 (Extended Data Fig. 8b, c). Downregulation of endogenous BRD4 decreased cell growth both in parental and resistant cells, which was rescued by enforced expression of wild-type BRD4 (Fig. 3c). BDmut BRD4 expression failed to rescue parental SUM159 cells, but supported growth of JQ1-resistant SUM159R consistent with an evident bromodomain-independent mechanism of BRD4 recruitment (Fig. 3c). Next, we assessed the sensitivity of cells

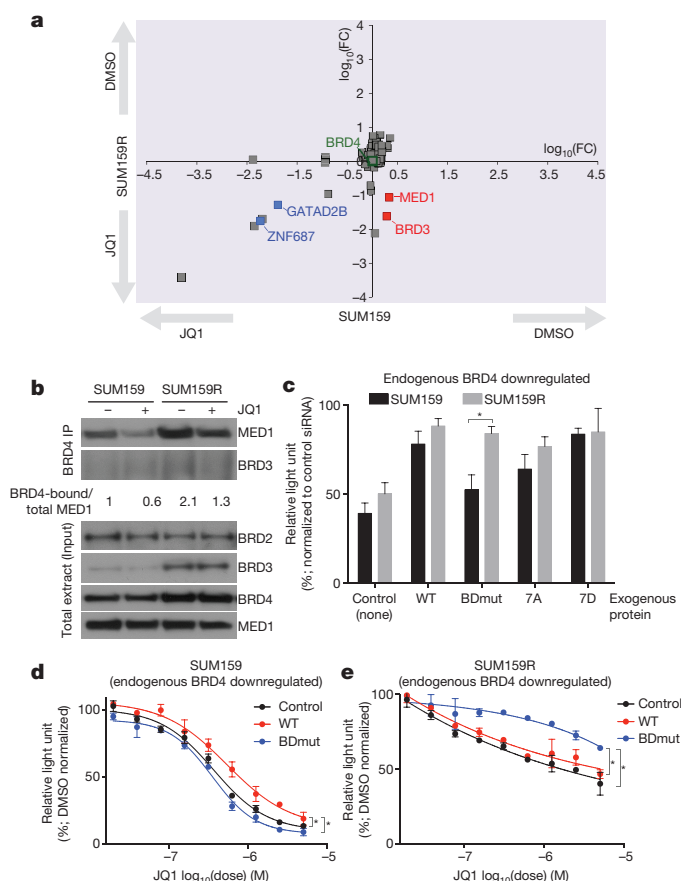


Figure 3 | Mechanism of BBI-resistance in TNBCs. All error bars represent s.d., $n = 3$. **a**, Plot depicting changes in BRD4-associated proteins in SUM159 and SUM159R cells following JQ1 treatment based on SILAC RIME. The axes represent \log_{10} of fold change (FC). **b**, Immunoblot analysis of BRD4 immunoprecipitates and total cell lysates in SUM159 and SUM159R cells. For gel source data, see Supplementary Fig. 1. **c**, Cellular viability of SUM159 and SUM159R cells expressing exogenous wild-type, BDmut, 7A and 7D mutant BRD4 with concomitant knockdown of endogenous BRD4. * $P = 0.04$ (paired *t*-test). **d**, **e**, Sensitivity of SUM159 (**d**) and SUM159R (**e**) cells expressing exogenous wild-type or BDmut BRD4 to JQ1 with concomitant knockdown of endogenous BRD4. * $P < 0.0001$ (non-linear regression, extra sum-of-squares test).

expressing BDmut BRD4 to JQ1, observing increased sensitivity to JQ1 in parental SUM159 cells exogenously expressing BDmut (Fig. 3d). In contrast, expression of BDmut BRD4 in SUM159R cells rescued the anti-proliferative effect of JQ1 (Fig. 3e), although this could partially be due to the slower growth of BDmut expressing cells. Together, these studies suggest that BBI-resistance is associated with increased binding of BRD4 to MED1, in a bromodomain-independent manner unaffected by JQ1.

A recent study reported that the stability and nuclear localization of BRD4 is increased with phosphorylation by casein kinase II (CK2)²¹. To explore the contribution of BRD4 phosphorylation to BBI-resistance, we performed immunoblot analysis in parental and resistant cells and found a marked increase of phospho-BRD4 (pBRD4) in resistant cells (Fig. 4a and Extended Data Fig. 8d). Small-molecule inhibition of CK2 decreased BRD4 phosphorylation in SUM159 and SUM159R cells (Extended Data Fig. 8e). These results imply BRD4 hyperphosphorylation in resistant cells either owing to increased phosphorylation by CK2 or, alternatively, to decreased dephosphorylation by an as yet unidentified BRD4 phosphatase. We therefore first analysed CK2 activity in parental and resistant cells by performing pan-CK2 substrate immunoblots and detected no significant differences in CK2 activity (Extended Data Fig. 8f).

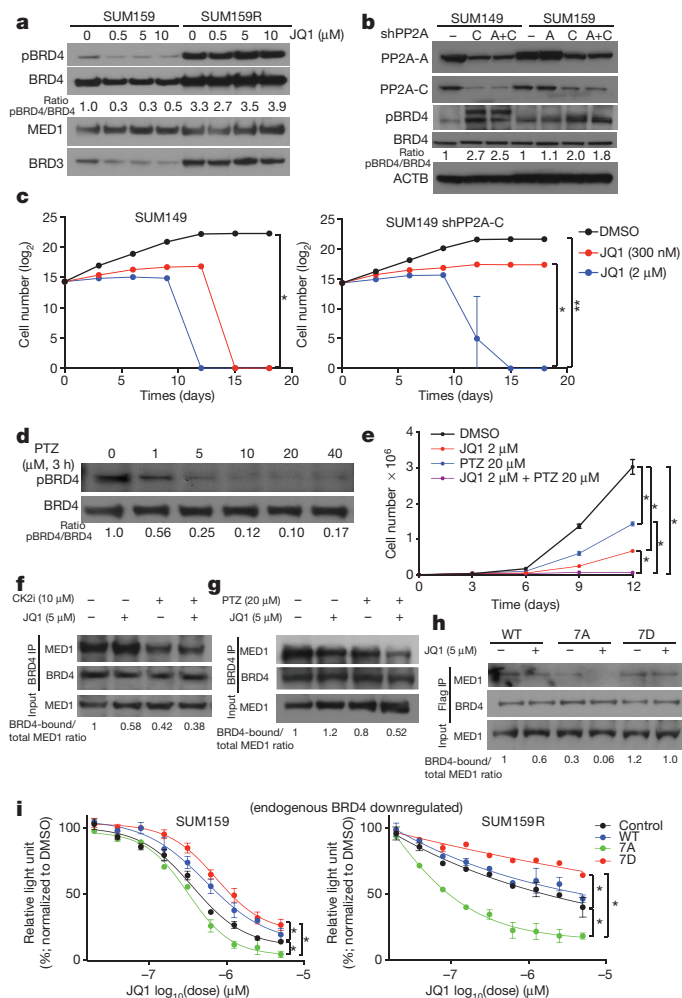


Figure 4 | Regulation and relevance of BRD4 phosphorylation.

All error bars represent s.d., $n = 3$. **a**, Immunoblot for the indicated proteins following JQ1 treatment. **b**, Immunoblot for the indicated proteins after knockdown of PP2A A or C or both subunits. **c**, Viable cell numbers of JQ1-treated control and shPP2A-C expressing SUM149 cells. Left, $*P < 0.0001$ (two-way ANOVA); right, $*P = 0.0037$ and $**P = 0.0015$ (two-way ANOVA). **d**, Immunoblot of pBRD4 and BRD4 in SUM159R cells following phenothiazine (PTZ) treatment. **e**, Viable cell numbers of SUM159R cells treated with JQ1, phenothiazine or both compounds. $*P < 0.0001$ (two-way ANOVA). **f, g**, Immunoblot of BRD4 immunoprecipitates and total cell lysates of SUM159R cells after 3 h treatment with JQ1 and CK2i (**f**) and JQ1 and PTZ (**g**). **h**, Immunoblot of Flag-BRD4 (wild type or mutant) immunoprecipitates and total cell lysates after 3 h treatment with JQ1. **i**, JQ1 sensitivity of SUM159 and SUM159R cells expressing exogenous wild-type or mutant (7A, 7D) BRD4. $*P < 0.0001$ (non-linear regression, extra sum-of-squares test). For gel source data, see Supplementary Fig. 1.

Inactivation of the PP2A phosphatase tumour suppressor gene occurs commonly in breast cancer and is associated with therapy resistance²²; PP2A also often opposes CK2 function^{23,24}. Thus, we investigated whether PP2A may dephosphorylate BRD4 and whether decreased PP2A activity could lead to BBI resistance. Downregulation of PP2A catalytic subunit (PPP2CA) in SUM149 and SUM159 cells led to increased BRD4 phosphorylation, establishing PP2A as a previously unrecognized BRD4 phosphatase (Fig. 4b), further supported by pharmacologic inhibitors of PP2A that showed similar effects (Extended Data Fig. 8g). To strengthen the link between PP2A activity and BBI resistance, we tested the JQ1 sensitivity of SUM149 cells following the knockdown of PP2A C subunit and determined that downregulation of PP2A decreased JQ1 sensitivity (Fig. 4c). We have collaboratively reported phenothiazine compounds as activators of PP2A enzymatic

activity²⁵. Thus, we analysed pBRD4 levels in SUM159R, SUM149R and other cell lines after short-term treatment with phenothiazine (PTZ) and detected rapid dephosphorylation of BRD4 (Fig. 4d and Extended Data Fig. 8h). Combined treatment with PTZ and JQ1 overcame BBI resistance in SUM159R cells (Fig. 4e). To investigate the functional role of BRD4 hyperphosphorylation in BBI resistance, we analysed whether BRD4 phosphorylation influences MED1 binding. Indeed, SUM159R cells treated with CK2 inhibitor or PTZ both lead to decreased MED1 abundance in BRD4 immunoprecipitations, suggesting that pBRD4 binds MED1 more efficiently than BRD4 (Fig. 4f, g).

To functionally assess the role for BRD4 phosphorylation in BBI resistance and MED1 binding, we generated BRD4 constructs encoding mutants that cannot be phosphorylated by CK2 (7 serine to alanine substitutions; '7A mutant') or mimic constitutive phosphorylation (7 serine to aspartate substitutions; '7D mutant'). We first assessed the ability of these constructs to rescue effects of endogenous BRD4 knockdown in stable cell lines (Extended Data Fig. 8b, c). We observed expression of both 7D and 7A mutants supporting the growth of both parental SUM159 and JQ1-resistant SUM159R cells (Fig. 3c). Next, we analysed MED1 binding and subcellular localization of 7A and 7D mutants in the presence or absence of JQ1. We found that the 7A mutant displays weaker MED1 binding compared to wild-type BRD4 and completely dissociates after JQ1, whereas the 7D mutant seems to have higher affinity for MED1 that is unaffected by JQ1 treatment (Fig. 4h and Extended Data Fig. 8i). Lastly, we assessed the sensitivity of cells expressing 7A or 7D mutant BRD4 to JQ1. In parental SUM159 cells exogenously expressed 7D mutant BRD4 decreased sensitivity to JQ1, whereas the 7A mutant slightly increased sensitivity (Fig. 4i). In contrast, expression of 7A mutant BRD4 in SUM159R cells restored JQ1 sensitivity, whereas the 7D mutant showed a modest decrease. These results strongly support the hypothesis that hyperphosphorylation of BRD4 arises from decreased PP2A activity in BBI resistant cells, leading to increased binding of BRD4 to MED1, recruitment to chromatin and decreased responsiveness to bromodomain inhibition.

To explore the clinical relevance of phospho-BRD4 (pBRD4) in BET inhibitor-naïve TNBC, we performed immunofluorescence analysis of a tissue microarray (TMA) featuring of 89 patient-derived TNBC specimens. First, we validated the pBRD4 immunofluorescence assay by comparing xenografts derived from SUM159 and SUM159R cell lines and detected significantly higher pBRD4 in SUM159R cells (Extended Data Fig. 9a). We detected strong pBRD4 staining among a subset of TNBCs (Extended Data Fig. 9b, c), and variable staining overall that was not correlated with expression of the androgen receptor and basal cytokeratins (bCK; Extended Data Fig. 9d, e) and it was not significantly associated with disease outcomes (Supplementary Table 9 and Extended Data Fig. 9f).

To extend the translational relevance of our findings, we conducted synergy studies of JQ1 with molecules targeting BCL-xL (ABT737), a gained super-enhancer in SUM159R cells, and modulators of BRD4 phosphorylation, the CK2 inhibitor CX-4945 and the PP2A activator perphenazine (PPZ). We observed significant synergy between JQ1 and all three compounds studied (Extended Data Fig. 10), establishing a rationale for combination studies of BBI in TNBC to improve response and to anticipate BBI resistance.

BRD4 inhibition has demonstrated efficacy in disparate models of cancer in a rapidly expanding literature. Despite apparent resistance in the vast majority of tumour types, as we observed here in TNBC, mechanisms of BBI-resistance have not been mechanistically explained. As this research was in review, two studies reported moderate emergent resistance to BBI in murine AML associated phenotypically with a stem-like state and WNT pathway activation^{26,27}. Interestingly, in our study TNBCs with more basal/stem cell-like features and WNT pathway activation are more sensitive to BET inhibition, whereas resistant disease emerges as epigenomic adaptation to a more differentiated luminal phenotype. Our findings of persistent BET bromodomain dependency despite BBI resistance, as well as pBRD4 staining in

resistant disease, should be studied in these murine AML models and further in human leukaemia.

Integrating approaches in epigenomics, proteomics and chemical biology, we provide an example of epigenomic drug resistance by an epigenetic mechanism, where in BBI-resistant cells, decreased PP2A activity leads to hyperphosphorylated BRD4, which binds more strongly to MED1, facilitating a bromodomain-independent chromatin recruitment mechanism. This research proposes putative combination strategies to anticipate and overcome BBI resistance, including pairing with BCL-xL inhibitors (for example, ABT-737) or CK2 inhibitors, and guides the development of second-generation BBIs that disrupt BET function via orthogonal biophysical or biochemical actions. More immediately, the robust efficacy observed in pre-clinical models supports the development of BET inhibition in TNBC alone, and in combination with mechanism-based targeted therapies.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 November 2014; accepted 3 December 2015.

Published online 6 January 2016.

- Vaz-Luis, I. *et al.* Outcomes by tumor subtype and treatment pattern in women with small, node-negative breast cancer: a multi-institutional study. *J. Clin. Oncol.* **32**, 2142–2150 (2014).
- Lehmann, B. D. *et al.* Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Invest.* **121**, 2750–2767 (2011).
- Metzger-Filho, O. *et al.* Dissecting the heterogeneity of triple-negative breast cancer. *J. Clin. Oncol.* **30**, 1879–1887 (2012).
- Puissant, A. *et al.* Targeting MYCN in neuroblastoma by BET bromodomain inhibition. *Cancer Discov.* **3**, 308–323 (2013).
- Delmore, J. E. *et al.* BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* **146**, 904–917 (2011).
- Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).
- Belkina, A. C. & Denis, G. V. BET domain co-regulators in obesity, inflammation and cancer. *Nature Rev. Cancer* **12**, 465–477 (2012).
- Chapuy, B. *et al.* Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* **24**, 777–790 (2013).
- Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
- Burstein, M. D. *et al.* Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clin. Cancer Res.* **21**, 1688–1698 (2015).
- Anders, L. *et al.* Genome-wide localization of small molecules. *Nature Biotechnol.* **32**, 92–96 (2014).
- Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
- Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
- Hayashi, H. *et al.* The *OCT4* pseudogene *POU5F1B* is amplified and promotes an aggressive phenotype in gastric cancer. *Oncogene* **34**, 199–208 (2015).
- Semenza, G. L. HIF-1 mediates metabolic responses to intratumoral hypoxia and oncogenic mutations. *J. Clin. Invest.* **123**, 3664–3671 (2013).
- Bessarabova, M. *et al.* Functional synergies yet distinct modulators affected by genetic alterations in common human cancers. *Cancer Res.* **71**, 3471–3481 (2011).
- Marotta, L. L. *et al.* The JAK2/STAT3 signaling pathway is required for growth of CD44⁺CD24[−] stem cell-like breast cancer cells in human tumors. *J. Clin. Invest.* **121**, 2723–2735 (2011).
- Ryan, J. & Letai, A. BH3 profiling in whole cells by fluorimeter or FACS. *Methods* **61**, 156–164 (2013).
- Montero, J. *et al.* Drug-induced death signaling strategy rapidly predicts cancer response to chemotherapy. *Cell* **160**, 977–989 (2015).
- Mohammed, H. *et al.* Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell Reports* **3**, 342–349 (2013).
- Wu, S. Y., Lee, A. Y., Lai, H. T., Zhang, H. & Chiang, C. M. Phospho switch triggers Brd4 chromatin binding and activator recruitment for gene-specific targeting. *Mol. Cell* **49**, 843–857 (2013).
- Rincón, R. *et al.* PP2A inhibition determines poor outcome and doxorubicin resistance in early breast cancer and its activation shows promising therapeutic effects. *Oncotarget* **6**, 4299–4314 (2015).
- Westermarck, J. & Hahn, W. C. Multiple pathways regulated by the tumor suppressor PP2A in transformation. *Trends Mol. Med.* **14**, 152–160 (2008).
- Eichhorn, P. J., Creighton, M. P. & Bernards, R. Protein phosphatase 2A regulatory subunits and cancer. *Biochim. Biophys. Acta* **1795**, 1–15 (2009).
- Gutierrez, A. *et al.* Phenothiazines induce PP2A-mediated apoptosis in T cell acute lymphoblastic leukemia. *J. Clin. Invest.* **124**, 644–655 (2014).
- Rathert, P. *et al.* Transcriptional plasticity promotes primary and acquired resistance to BET inhibition. *Nature* **525**, 543–547 (2015).
- Fong, C. Y. *et al.* BET inhibitor resistance emerges from leukaemia stem cells. *Nature* **525**, 538–542 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Silver and members of the Polyak and Bradner laboratories for their critical reading of this manuscript and useful discussions. We thank G. Brown for help with creating the word cloud figures. This work was supported by the NIH DF/HCC SPORE in Breast Cancer CA168504 (K.P., E.P.W., I.E.K., D.D., W.T.B., and J.E.B.), CA080111 (K.P. and M.B.), and CA103867 (C.M.C.), Susan G. Komen Foundation (S.S.), CPRIT RP110471 and RP140367 (C.M.C.), Welch Foundation (C.M.C.), US Department of Defense CDMRP BC122003 (S.X.L.) and CA120184 (C.Y.L.), Princess Margaret Cancer Foundation (H.H.H.), Canada Foundation for Innovation and Ontario Research Fund CF132372 (H.H.H.), NSERC discovery grant RGPIN-2015-04658 (H.H.H.), and the Harvard Ludwig Center for Cancer Research (J.E.B., M.B. and K.P.).

Author Contributions S.S. performed cell culture, xenograft, ChIP-seq, and RNA-seq experiments, and data analyses. C.Y.L. and C.A.M. performed genomic data analyses. H.H.H. helped with ChIP-seq and RNA-seq experiments and data analyses. R.M.W. performed cell culture, ChIP-seq experiments and data analyses. J.M.R. performed synergy studies. D.P.T. helped with immunofluorescence staining. M.J. and S.J.H. helped with confocal microscopy and image quantification. Y.L. helped with BRD4 ChIP-seq. M.B.E. and G.P. helped with cell cycle studies. E.D. helped with generating and testing BRD4 mutants. J.B. and L.A. performed Chem-seq. H.M., C.D. and J.S.C. conducted proteomic experiments and data analyses. C.O. and M.M. performed drug sensitivity screens. J.Q. synthesized BBI compounds. M.N. generated shRNA constructs. D.D., I.E.K. and E.P.W. generated the TMA and linked to clinical data. H.G., D.G.S. and W.T.B. performed TMA and statistical analyses. J.R. and A.L. performed BH3 profiling and data analyses. C.-M.C. and S.-Y.W. provided phospho-BRD4 antibody. P.K.R. and M.D. generated RNA-seq libraries. K.P. supervised with help from J.E.B., X.S.L., M.B., R.A.Y. and H.L. All authors helped to design the study and write the manuscript.

Author Information RNA-seq, ChIP-seq, and Chem-seq data have been deposited in the NCBI GEO database with the accession number GSE63584. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.P. (kornelia_polyak@dfci.harvard.edu) or J.E.B. (james_bradner@dfci.harvard.edu).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

Cell lines and breast tumour tissues. Breast cell lines were obtained from the ATCC and S. Ethier (SUM series). Cells were cultured in media recommended by the provider, their identity was confirmed by STR analysis, and they were regularly tested for mycoplasma. Breast tumour samples were collected using protocols approved by the DF/HCC Institutional Review Board, informed consent was obtained from all patients. Tumours were minced with razor blades and digested with stirring for 3–4 h at 37 °C in DMEM/F12 with 2 mg ml⁻¹ BSA, 2 mg ml⁻¹ collagenase type IV, and 2 mg ml⁻¹ hyaluronidase. After digestion, cells were filtered through 500-µm mesh, washed in DMEM/F12 with 5% FBS, frozen in DMEM/F12 with 5% FBS and 10% DMSO, and stored in liquid nitrogen for subsequent xenograft studies. PDX IDC50 was derived from a primary tumour of highly invasive metaplastic TNBC resistant to chemo and radiation therapy leading to the rapid death of the patient. Exome sequencing of the tumour and xenograft identified numerous mutations including heterozygous frameshift mutation in *PTEN* (chr10_89701964-89701964_A) and *CDH1* chr16_67400242-67400242_C). PDX EL-12-58 was derived from a liver metastasis of a heavily pretreated basal-like TNBC, Oncopanel mutation testing identified homozygous mutations in *BRCA2* (p.S1970*), *TP53* (p.I232fs), *TSC2*, *FLT3*, and *ROS1*, and lower frequency mutations in *RAD21*, *JAK3*, *ARID1B*, *ARID1A*, *KDM6A*.

High-throughput screening of BET bromodomain inhibitors in breast cell line panel. We tested a panel of compounds (synthesized in the Bradner laboratory) in 40 human breast cell lines in a 384-well format at 2,000 cells per well using a semi-automated screen essentially as described⁵. Cell viability at 72 h was evaluated using ATPlite (Perkin Elmer).

Synergy studies. SUM149, SUM149R, SUM159, and SUM159R cells were seeded in sterile, white, opaque 384-well microtitre plates (Thermo), using an automated dispensing system (BioTek EL406), at 1,000 cells per well in 50 µl of media. Drugs were delivered in DMSO by robotic pin transfer with a JANUS workstation (100 nl) to achieve a matrix of pairwise dose–response incubations of each compound, each pair having eight replicates. Following 72 h of incubation, ATP levels were determined for treated cells and vehicle controls (ATPlite, PerkinElmer). Data were normalized to vehicle controls. Combination indices were determined using the median-effect principle of Chou & Talalay²⁸ (CalcuSyn Software). Isobologram plots were generated with GraphPad Prism software. Points represent paired values of drug concentrations assessed for synergism. The diagonal line signifies drug additivity. Points above the line represent antagonistic drug combinations, and those below the line represent synergistic drug combinations. Synergy assays were performed in triplicates and repeated 2–3 times.

Xenograft assays. For xenograft assays 5–6-weeks old female CrTac:NCr-Foxn1^{tmu} and NOD.Cg-Prkdc^{scid} IL2rg^{tm1Sug}/JicTac mice were purchased from Taconic. Tumours were induced by bilateral orthotopic mammary fat pad injection of 1 × 10⁶ cells in 50% Matrigel (BD Biosciences) in DMEM/F12 or Medium 171 (except for IDC50-X cells, which were injected with 3% FBS and 4 mg ml⁻¹ collagen gel in Medium 171). Animal experiments were conducted following protocol 11-023 approved by the Dana-Farber Cancer Institute Animal Care and Use Committee. For all the xenograft studies, the sample size of each group (5–10 mice) is indicated in the figures. We performed pilot experiments using a few (5–10) mice per group followed by larger studies if needed to reach statistical significance and repeated experiments to ensure reproducibility. Due to the nature of the performed experiments, no randomization and no blinding was used as it was deemed unfeasible. However, the resulting tumours were analysed in a blinded manner. Mice were administered JQ1 (50 mg per kg, daily), vehicle only (control) for 14 days beginning at day 14 (SUM159), or doxycycline at day 21 (SUM159-shBRD4) after injection. Mice were euthanized and tumours evaluated 28 and 60 days after injection of parental and TET-inducible shBRD4-expressing SUM159 cells into mammary fat pads.

Cellular viability, senescence, MDR and BH3 profiling assays. Cell viability and growth assays (Figs 1a, 2a, b, 3d, e, 4c, i, Extended Data Figs 1a, b, 3i, 4d, e, g, h, 10), cycle, apoptosis, and MDR assays were performed in triplicates and repeated 2–3 times. For cell proliferation assays, cells were plated at 500 cells per well in 96-well plates and treated the next day with inhibitors, DMSO or doxycycline (500 ng). Cells were cultured at 37 °C with 5% CO₂ in the media described above, and cell viability was measured using CellTiter-Glo three days after treatments. For cell growth assays, cells were plated at 5,000 (SUM159) or 20,000 (SUM149) cells per well in 6-well plates and treated the next day with inhibitors. Cells were counted every three days by cell counter. Cellular apoptosis was analysed with an APC AnnexinV/7AAD Apoptosis Detection kit (BD Pharmingen). AnnexinV/7AAD assessments and cell cycle graphics were generated using FlowJo software V7.6.1 for Windows (Tree Star). Senescence Beta-gal staining was performed using

Senescence β-Galactosidase staining kit from Cell Signaling. Briefly, after JQ1 treatment (500 nM) for 72 h, SUM159 and MDA-MB-231 cells were fixed by fixative solution for 15 min, followed by β-galactosidase solution incubation overnight at 37 °C. The staining was checked under microscope for the development of blue colour. Multi-Drug Resistance Assay was performed with MDR assay kit from Cayman Chemical (600370). Briefly, SUM159 and SUM159R cells were treated with JQ1 or DMSO for 30 min in SUM medium. Verapamil was used as a positive control at 1:1,000 dilution. Calcein AM/Hoechst Dye staining solution was added after that and cells were incubated at 37 °C for 15 min. The cells were analysed by fluorescent microscope and FACS.

Cell cycle analysis was performed 72 h after JQ1 treatment or BRD4 down-regulation with doxycycline using propidium iodide (PI) staining. Cells were resuspended in 1 ml of growth medium supplemented with 2 µg ml⁻¹ PI (Life Technologies) as final concentration. After 60 min at 37 °C in the dark, analysis was performed on a FACS AriaII cytometer (BD Biosciences). The cell cycle was plotted as histogram after excluding doublets.

Cell synchronization procedure. SUM159 cells were treated with nocodazole (200 ng ml⁻¹) for 12 h and then cells were tapped to detach from the plates. After washing twice with PBS, cells were replated with or without JQ1 in collagen-coated plates. Cells were collected at 0, 3, 6, 12 h time point for FACS and immunoblot analysis. Dynamic BH3 profiling was performed using the JC-1 plate method as previously described^{18,19}. Briefly, 2.5 × 10⁵ cells were seeded in T25 flasks in the presence of 500 nM or 5 µM JQ1 for 72 or 96 h. Cells were trypsinized, suspended in MEB (150 mM mannitol, 10 mM HEPES, 50 mM KCl, 5 mM succinate, 20 µM EDTA, 20 µM EGTA, 0.1% protease-free BSA, pH 7.5 ± 0.1), and 1–2 × 10⁴ cells were added in 15 µl of MEB to each well of a 384 well Fluotrac 200 plate containing 15 µl per well of either peptides at 2 × final concentration, buffer only, or 50 µM alamethicin in MEB supplemented with 2 µM JC-1, 10 mM 2-mercaptoethanol, 20 µg ml⁻¹ oligomycin, and 50 µg ml⁻¹ digitonin. Fluorescence at emission 590 ± 10 nM and excitation 545 ± 10 nM was recorded at 5 min intervals at 32 °C. The area under each curve was normalized to the alamethicin and buffer controls as:

$$\% \text{ Depolarization} = 1 - [(AUC_{\text{sample}} - AUC_{\text{alamethicin}}) / (AUC_{\text{buffer}} - AUC_{\text{alamethicin}})]$$

Delta priming was calculated per peptide treatment as: Delta Priming = (Depolarization Treated) – (Depolarization Untreated). Positive delta priming indicates an increase in priming due to treatment and an increased potential for apoptosis at later time points.

Immunofluorescence staining and image and statistical analysis of tissue microarrays. Antibodies used for immunofluorescence were: CK18 (Dako, M7010), CK17 (Dako, M7046), HMW CK (Dako, M0630), LMW CK (Dako, M0631), CD44 (NeoMarkers, MS-668-P1), CD24 (NeoMarkers, MS-1279-P1), pSTAT3 (Cell Signaling, 9145S), VIM (Dako, M073501), CDH1 (BD Biosciences, 610181), Flag (Sigma, F1804), BrdU (Roche, 11170376001), pBRD4 (a gift from C. M. Chiang), and androgen receptor (Cell Signaling, 5153S). Immunofluorescence experiments were performed in cultured cells or in whole sections of formalin-fixed paraffin embedded (FFPE) xenograft tumours. The staining was performed as described²⁹. Antibody dilutions were as follows: pSTAT3 (1:25), CD44 (1:100), CD24 (1:100), CK18 (1:200), CK17 (1:200), HMW CK (1:100), LMW CK (1:100), VIM (1:100), CDH1 (1:100), Flag (1:50), BrdU (1:200), pBRD4 (1:200), and androgen receptor (1:50). The Dana-Farber Breast Cancer Tissue Microarray (TMA) consists of primary TNBC samples from approximately 83 patients who underwent definitive breast surgery at Brigham and Women's Hospital between 1 January 1997 and 31 December 2005. Formalin-fixed, paraffin-embedded breast cancers were collected from the archives of the Department of Pathology at Brigham and Women's Hospital. Best blocks and best areas for coring were identified and selected by a breast pathologist (DD) to represent different area of the tumour. Results of immunohistochemical studies for oestrogen (ER) and progesterone receptor (PR) and HER2 and FISH assay results for HER2 were extracted from pathology reports. TMA construction was carried out in the Dana-Farber/Harvard Cancer Center Tissue Microarray Core Facility. Three 0.6 mm cores were taken from different marked areas in each case and placed into a recipient block using a manual arrayer (Beecher Instruments). Specimens are arrayed in triplicate. Participants signed consent for research use of tissue and the linking of tumour specimens to clinical follow-up. Clinical data on these patients was collected retrospectively at first presentation, at 4, 9, 18, 30, and 42 months, and annually thereafter. After 9.3 years median follow up, 24 recurrences and 14 deaths have been recorded. The data elements are the following: staging, tumour pathology, diagnostic and follow-up tests performed, treatments administered (surgery, radiation and systemic therapy), and recurrence. Although the patients in this cohort were not treated as part of a clinical trial protocol, they were treated relatively uniformly as per Dana-Farber clinical practice guidelines. This serves to minimize confounding due to treatment heterogeneity. The TMA was stained with pBRD4 (1:200) antibody and imaged manually on Yokagawa spinning disc confocal microscope.

Three images were taken per each core for 240 out of 267 cores, for the remaining 27 one or two images were taken due to tissue loss or low tumour content. Image analysis was performed with ImageJ software macro (code available upon request). Phospho-BRD4 staining mean intensity was calculated per individual nucleus within an image. The mean intensity per image was normalized to nuclei count. For clinical outcome analysis patients were dichotomized as 'High'/'Low' pBRD4 by median intensity (Supplementary Table 9). Disease outcomes were evaluated in 83 of 89 TMA samples (3 were not TNBC by definitive pathology, 2 did not have clinical data available, one was a repeat biopsy on a patient). Recurrence-free survival (RFS) was defined as the interval from the date of initial surgical resection to the date of recurrence (local or distant), or date of last known contact if the patient was alive and has not recurred. RFS and overall survival were estimated using the Kaplan–Meier product-limit method, with hazard ratios and 95% confidence intervals from a univariate Cox proportional hazard model.

siRNAs and lentiviral shRNA and expression constructs. For siRNA transfection cells were plated at 2,000 cells per well in 96-well plates and cultured at 37 °C with 5% CO₂ in the media. The next day, cells were transfected in triplicate with siGENOME SMARTpools for the genes of interest or "Non-Targeting siRNA" controls using DharmaFECT 1 (Dharmacon). The sequences of the siRNAs in the SMARTpools are listed in Supplementary Table 10. Cell viability was measured using CellTiter-Glo (Promega) three days after transfections, with the effects of each siRNAs treatment on each cell line compared to the effects of no siRNAs.

TET-inducible pLKO-TET-ON lentiviral constructs were packaged by co-transfection of the lentiviral hairpin containing plasmid pLKO.1 and the helper plasmids pCMV-dR8.91 and pMD2.G-VS.V-G into HEK293T cells using Lipofectamine (Life Technologies). Following transduction via spinoculation for 30 min at 1,000g and selection with 1 µg ml⁻¹ puromycin for 72 h (Sigma, St. Louis, MO), knockdown efficacy was determined by western blotting and cells were seeded for proliferation assays as described above. Sequences of shRNAs used are listed in Supplementary Table 10.

Full length BRD4 in pCDNA3 was a gift from Dr. French at Brigham and Women's Hospital, Harvard Medical School. Mutations of BRD4 BD1 (N140A) and BD2 (N433A) bromodomains, 7A and 7D mutants were generated using a Quickchange Multi Site-Directed Mutagenesis Kit (Agilent Technologies) using primers listed in Supplementary Table 10 and subsequently verified by sequencing. **Immunoblotting and immunoprecipitation experiments.** Cells were lysed five days after transfection with siRNAs in RIPA buffer. Proteins were resolved in SDS-polyacrylamide gels (4–12%) and transferred to PVDF membranes by using a Tris-glycine buffer system. Membranes were blocked with 5% milk powder in 0.1% Tween20 in PBS (PBS-T) for 1 h at room temperature followed by incubation with primary antibodies at 1:1,000 dilution in 2.5% milk PBS-T. For immunoprecipitation, nuclear extracts were prepared as follow: 1×10^7 cells were resuspended in 5 ml buffer A: 10 mM Tris pH 7.9, 1.5 mM MgCl₂, 10 mM KCl, 0.05% NP-40, 1 mM DTT, and protease and phosphatase inhibitors. Cells were incubated on ice for 15 min and gently vortexed every 5 min. After centrifugation at 2,000g for 5 min, pellets were suspended in 0.3 ml buffer B (20 mM Tris pH 7.9, 25% glycerol, 0.42 M NaCl, 1.5 mM MgCl₂, 1 mM KCl, 0.5% NP40, 0.2 mM EDTA, 1 mM DTT, and protease and phosphatase inhibitors) and incubated for 5 min on ice. After centrifugation of the lysates at 14,000g for 10 min at 4 °C, supernatant was diluted with 0.6 ml buffer A, and added NP-40 to final 0.5% and treated with DNase I. The samples were then incubated at 4 °C overnight with BRD4 or Flag antibodies at 1:100 dilution and immunoprecipitates were collected on Dynabeads Protein G for 2 h. Beads were washed with buffer B containing 150 mM NaCl and 0.5% NP-40 three times and then resuspended in gel loading buffer. Immunoblotting and immunoprecipitation experiments were repeated 2–3 times.

Antibodies and inhibitors. Antibodies used for immunoblotting, immunoprecipitation and ChIP-seq were as follows: BRD4 (Bethyl, A301-985A), MED1 (Bethyl, A300-793a), BRD3 (Bethyl, A302-368A), BRD2 (Bethyl, A302-583A), MYC (Santa Cruz, sc764), pSTAT3 (Cell Signaling, 9145S), STAT3 (Cell Signaling, 4904), pSTAT5 (Cell Signaling, 9351), pJAK2 (Cell Signaling, 3771), CYCLIN D1 (Cell Signaling, 2922), pH3 (Cell Signaling, 12201), CK2 substrate (Cell Signaling, 8738), PP2A-A (Cell Signaling, 2039), PP2A-C (Cell Signaling, 2038) and pBRD4 was a gift from C. M. Chiang. Antibodies used for ChIP-seq were BRD4 (Bethyl) histone H3K27ac (Abcam, ab4729), CXCR2 inhibitor (239819) and CK2 inhibitor (218860) were from CalBiochem, JAK2 inhibitor (Ruxolitinib, INCB018424), MEK inhibitor (GSK1120212, S2673), ABT-737 (s1002), and PI3K inhibitor (BKM120, S2247) were from Selleckchem, phenothiazine (1525707) and perphenazine (1511000) were from Sigma-Aldrich. Inhibitor treatment for immunoblot analyses was conducted for 3 h.

SILAC-RIME experiments and data analysis. SUM159 and SUM159R cells were grown in R/K-deficient SILAC DMEM (paa; E15-086), 10% dialysed serum (Sigma-Aldrich; F0392), and supplemented with 800 µM L-lysine ¹³C₆¹⁵N₂

hydrochloride and 482 µM L-arginine ¹³C₆¹⁵N₄ hydrochloride (Cambridge Isotope laboratory) for 'heavy'-labelled media or 800 µM L-lysine ¹²C₆¹⁴N₂-hydrochloride and 482 µM L-arginine ¹²C₆¹⁴N₄ hydrochloride for 'light'-labelled media. After SILAC labelling, RIME was performed as described²⁰. Word clouds for Extended Data Fig. 7 were generated using R version 3.1.0 and the R package 'wordcloud' version 2.5. The size of the tag reflects the square root of the MASCOT score of the protein (the choice of square root is arbitrary, but visually appealing). Experiments were filtered against the Contaminant Repository for Affinity Purification Mass Spectrometry Data³⁰, considering any protein which occurs in at least 20 negative control experiments to be contamination, hence removed from the data set. Refseq protein IDs provided by the contaminant repository were converted to Uniprot IDs found in the mass spec experiments using mappings from the Bioconductor package 'org.Hs.eg.db', version 2.14 (Carlson M. org.Hs.eg.db: Genome wide annotation for Human. R package version 3.0.0). SILAC RIME experiments were performed in duplicates and repeated 2–3 times.

In vitro Chem-seq, ChIP-seq, and RNA-seq. Chem-seq was performed essentially as described¹¹. BRD4 ChIP-seq: SUM159 and SUM159R cells (4×10^7) were grown in SUM Medium. The media were then removed and replaced with media containing 1% formaldehyde (EM grade; tebu-bio) and crosslinked for 8 min. Crosslinking was quenched by adding glycine to a final concentration of 0.2 M. The cells were washed with ice-cold PBS, harvested in PBS, and the cell pellet was washed with PBS. The nuclear fraction was extracted by first resuspending the pellet in 10 ml of LB1 buffer (50 mM HEPES-KOH (pH 7.5), 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40 or Igepal CA-630, and 0.25% Triton X-100) for 10 min at 4 °C. Cells were pelleted, resuspended in 10 ml of LB2 buffer (10 mM Tris-HCL (pH 8.0), 200 mM NaCl, 1 mM EDTA, and 0.5 mM EGTA), and mixed for 5 min. Cells were pelleted and resuspended in 300 µl of LB3 buffer (10 mM Tris-HCL (pH 8), 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-deoxycholate, and 0.5% N-lauroylsarcosine) and sonicated in a Covaris sonicator for 10 min. A total of 30 µl of 10% Triton X-100 was added, and lysate was centrifuged for 10 min at 20,000 rcf to purify the debris. The supernatant was then incubated with 100 µl of Dynabeads Protein G (Life Technologies, 10003D) prebound with 20 µg BRD4 antibody (Bethyl, A301-985A), and immunoprecipitation (IP) was conducted overnight in the cold room. The beads were washed ten times in 1 ml of RIPA buffer and twice in 100 mM ammonium hydrogen carbonate (AMBI) solution. DNA was eluted in elution buffer (50 mM Tris-HCL pH 8, 10 mM EDTA, and 1% SDS). Cross-links were reversed overnight at 65 °C. RNA and protein were digested with 0.2 mg ml⁻¹ RNase A for 2 h followed by 0.2 mg ml⁻¹ Proteinase K for 1 h. DNA was purified with phenol-chloroform extraction and ethanol precipitation. Libraries for Illumina sequencing were prepared following the Rubicon ThruPLEX-FD kit for 10–12 cycles.

RNA-seq: SUM159 and SUM159R were incubated in biological duplicates for 3, 12 and 24 h with 500 nM of JQ1 or DMSO treatment. Total RNA was extracted using the standard Qiagen RNeasy kit (74106). RNA concentrations were measured and quality controlled on a Bioanalyzer, RNA-Seq libraries were made using Illumina True-Seq RNA kits using the Sciclone NGSx workstation.

All RNA-seq and ChIP-seq experiments were performed in duplicates.

Genomic data analyses. Accessing data generated in this manuscript. All ChIP-seq, Chem-seq, and RNA-seq data generated in this publication can be found online associated with GEO Publication Reference ID GSE63584 (www.ncbi.nlm.nih.gov/geo/). Supplementary Table 2 lists all sequencing data sets and their corresponding GEO GSM accession IDs.

Gene sets and annotations. All analysis was performed using RefSeq (NCBI37/HG19) human gene annotations.

RNA-seq data processing and gene expression quantification. All RNA-Seq data sets were aligned to the transcriptome using Tophat2³¹ (version 2.0.11) using the Illumina genomes NCBI37/HG19 UCSC transcriptome build retrieved from <http://ccb.jhu.edu/software/tophat/igenomes.shtml>. Alignments were performed using default parameters. Transcript expression quantification was performed using Cufflinks³² (version 2.2.0) with default parameters to generate gene expression values in units of FPKM.

ChIP-seq and Chem-seq data processing. All ChIP-seq and Chem-seq data sets were aligned using Bowtie2³³ (version 2.2.1) to build version NCBI37/HG19 of the human genome or build version NCB37/MM9 of the mouse genome. Alignments were performed using the following criteria: -k 1, with all other parameters set to default. These criteria preserved only reads that mapped uniquely to the genome without any mismatches.

Calculating read density. We calculated the normalized read density of a ChIP-seq or Chem-seq data set in any region using the Bamliquidator (version 0.9) read density calculator (<https://github.com/BradnerLab/pipeline/wiki/bamliquidator>). Briefly, ChIP-Seq reads aligning to the region were extended by 200 bp and the density of reads per base pair (bp) was calculated. The density of reads in each region was normalized to the total number of million mapped

reads producing read density in units of reads per million mapped reads per bp (rpm per bp).

Identifying ChIP-seq and Chem-seq enriched regions. We used the MACS version 1.4.2 (Model based analysis of ChIP-Seq)³⁴ peak finding algorithm to identify regions of ChIP-Seq enrichment over background. A *P* value threshold of enrichment of 1×10^{-9} was used for all data sets. The GEO accession number and background used for each data set can be found in the accompanying Supplementary Table 2.

Creating heat map representations of ChIP-seq occupancy. Heat maps of ChIP-seq occupancy for various factors were created as described³⁵. Heat maps were created for the ± 10 kb region flanking all transcription start sites (TSS) or for the ± 10 kb region flanking all TSS distal BET bromodomain bound enhancers. Each row plots a specific TSS or enhancer region. Rows are ranked by peak occupancy of BET bromodomains as determined by Bio-JQ1 Chem-seq signal (Fig. 1d).

Correlating BRD4 and H3K27ac occupancy to Bio-JQ1. Occupancy of BRD4 and H3K27ac was correlated to Bio-JQ1 occupancy at all regions of Bio-JQ1 enrichment in SUM159 cells. Pearson correlation statistics are shown (Extended Data Fig. 3a). To quantify changes in BRD4 or H3K27ac occupancy upon JQ1 treatment, all Bio-JQ1 enriched regions were ranked in SUM159 cells and then binned ($n = 10$). Corresponding box plots of BRD4 or H3K27ac \log_2 fold change with or without JQ1 are shown for each bin (Extended Data Fig. 3b).

Mapping enhancers and super-enhancers using Bio-JQ1 occupancy or BRD4. Enhancers and super enhancers (SEs) were mapped using the ROSE software package described^{12,13} and available at (http://younglab.wi.mit.edu/super_enhancer_code.html). In SUM159 and SUM159R cells, Bio-JQ1 Chem-Seq enriched regions were used to map enhancers and super enhancers (Fig. 1f). In SUM149 cells, BRD4 ChIP-seq enriched regions were used to map enhancers and super enhancers (Extended Data Fig. 3a). Enhancers are defined as regions of Bio-JQ1 binding not contained in promoters.

Quantifying changes in gene expression of super enhancer proximal genes. Genes within 50 kb of super enhancer in SUM159 or SUM149 were identified and filtered for expression status (>1 FPKM expression in any sample), and filtered to remove non poly-adenylated transcripts (for example, microRNAs). For SUM159, \log_2 fold changes in gene expression at super-enhancer-associated genes or all expressed genes was compared at 3, 12, and 24 h post JQ1 treatment (Fig. 1g). For SUM149, comparisons were made at 12 h post JQ1 treatment (Extended Data Fig. 3f). The statistical significance of differences between distributions of changes was also assessed using a Welch's two-tailed *t* test.

Identifying differentially expressed genes upon JQ1 treatment. To identify genes differentially regulated by JQ1 treatment in SUM159 or SUM149 cells, all genes with a $>1 \log_2$ fold change in expression were ordered by fold change at 24 h with or without JQ1 for SUM159 or at 12 h with or without JQ1 for SUM149. The \log_2 row median normalized fold change for each gene is displayed as a heat map in Extended Data Fig. 3g for SUM159 and in Extended Data Fig. 3h for SUM149. For subsequent gene set and pathway analysis, SUM159 genes with consistent and statistically significantly altered expression were selected using a Welch's two-tailed *t* test between DMSO and JQ1 treated expression values at 12 and 24 h. A *P* value cut-off of 0.01 was applied (Extended Data Fig. 3k).

Identifying gained/lost super enhancers between SUM159 and SUM159R. Super enhancer differential regions were defined as in Brown *et al.* 2014³⁶. Briefly, in order to quantify changes in super-enhancers between two conditions, background subtracted ChIP-Seq signal was calculated at the set of all enhancer regions considered super in at least one condition. Gained/lost super-enhancers were determined as those with a greater than \log_2 fold change signal in either direction. The \log_2 fold change in Bio-JQ1 occupancy at all rank ordered super-enhancer-containing regions is shown in Fig. 2c. Super enhancer regions were classified as either gained, conserved, or lost. Gained/lost regions were classified as those with $>1 \log_2$ fold change in either direction. Conserved regions were classified as those with $<0.25 \log_2$ fold change in either direction. The \log_2 fold change in either BRD4 or proximal (within 50 kb of region) gene expression is shown in Extended Data Fig. 5d–f.

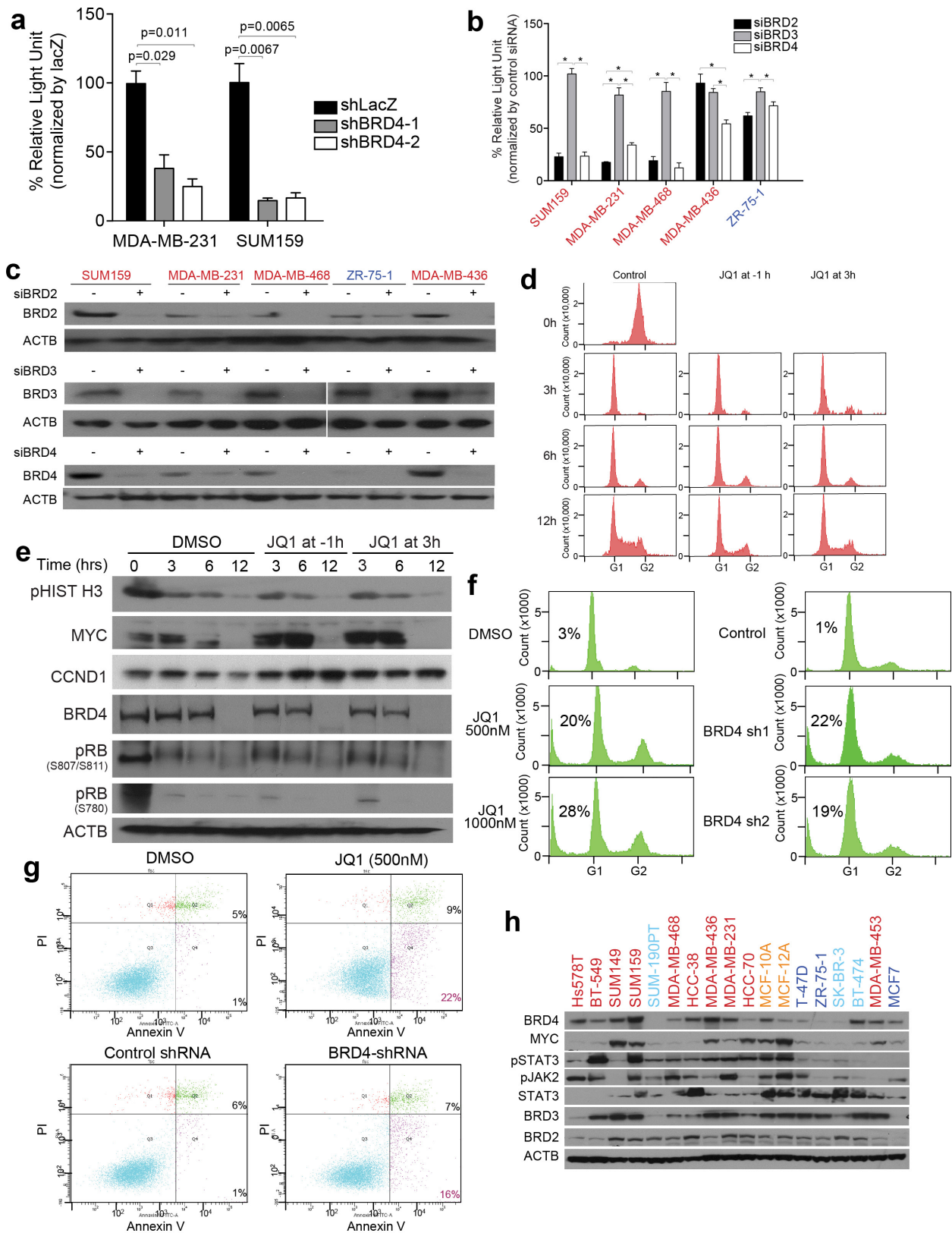
Quantifying changes in BRD4 and H3K27ac occupancy upon JQ1 treatment in either SUM159 or SUM159R cells at Bio-JQ1 regions. \log_2 fold changes in BRD4 or H3K27ac were quantified at Bio-JQ1 enriched regions in their respective cell line and shown in Extended Data Fig. 5c.

Quantifying changes in BRD4 and H3K27ac as a function of Bio-JQ1 or BRD4 occupancy. Bio-JQ1 enriched regions in SUM159 or BRD4 enriched regions in SUM149 were ranked by increasing levels and then distributed into 10 bins. \log_2 fold changes in BRD4 or H3K27ac were quantified in each bin of regions and displayed as a box plot (Extended Data Fig. 3b, c).

Quantifying changes in BRD4 occupancy upon JQ1 treatment in all TNBC. \log_2 fold changes in BRD4 upon JQ1 treatment were quantified at BRD4 enriched regions in each respective cell line Extended Data Fig. 6c.

All code related to genomic and transcriptome analysis can be found at <https://github.com/BradnerLab/TNBC>.

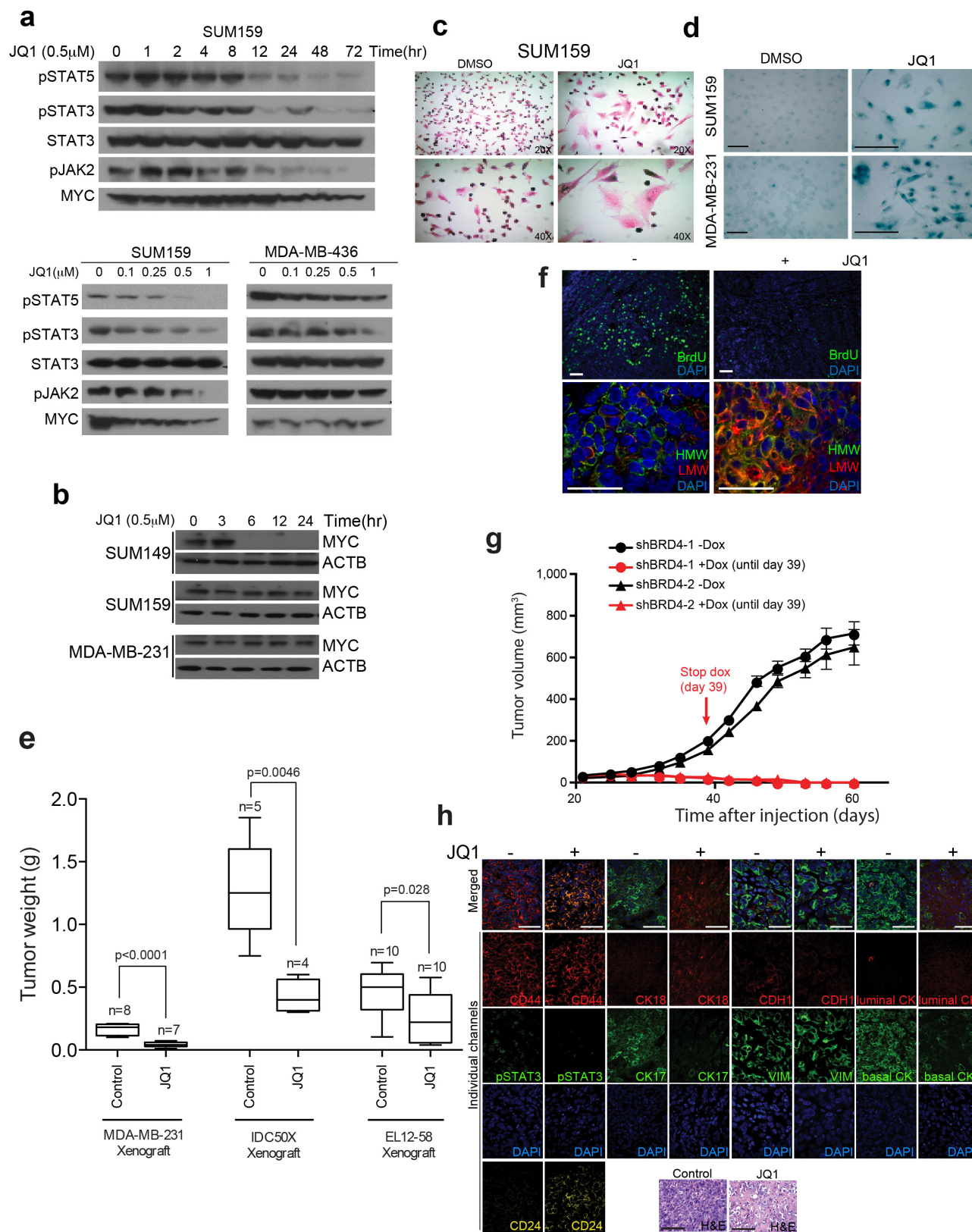
28. Chou, T. C. & Talalay, P. Quantitative analysis of dose-effect relationships: the combined effects of multiple drugs or enzyme inhibitors. *Adv. Enzyme Regul.* **22**, 27–55 (1984).
29. Choudhury, S. *et al.* Molecular profiling of human mammary gland links breast cancer risk to a p27⁺ cell population with progenitor characteristics. *Cell Stem Cell* **13**, 117–130 (2013).
30. Mellacheruvu, D. *et al.* The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nature Methods* **10**, 730–736 (2013).
31. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
32. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28**, 511–515 (2010).
33. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
34. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
35. Lin, C. Y. *et al.* Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **151**, 56–67 (2012).
36. Brown, J. D. *et al.* NF- κ B directs dynamic super enhancer formation in inflammation and atherogenesis. *Mol. Cell* **56**, 219–231 (2014).



Extended Data Figure 1 | See next page for figure caption.

Extended Data Figure 1 | BET bromodomain proteins and cell growth in TNBCs. All error bars represent s.d., $n = 3$. **a**, Cellular viability of SUM159 and MDA-MB-231 cells expressing TET-inducible BRD4-targeting or lacZ shRNAs. P values indicate statistical significance of the observed differences (paired t -test). **b**, Cellular viability four days after transfection of siRNAs targeting BET bromodomain proteins. Asterisks indicate statistical significance (paired t -test) of the marked differences as follows. SUM159: siBRD2 versus siBRD3, $P = 0.002$; siBRD3 versus siBRD4, $P = 0.0006$, MDA-MB-231: siBRD2 versus siBRD3, $P = 0.006$; siBRD2 versus siBRD4, $P = 0.002$; siBRD3 versus siBRD4, $P = 0.016$, MDA-MB-468: siBRD2 versus siBRD3, $P = 0.0009$; siBRD3 versus siBRD4, $P = 0.0055$, MDA-MB-436: siBRD2 versus siBRD4, $P = 0.002$; siBRD3 versus siBRD4, $P = 0.015$, ZR-75-1: siBRD2 versus siBRD3, $P = 0.0169$; siBRD3 versus siBRD4, $P = 0.007$. **c**, Immunoblot analysis of BET bromodomain proteins four days after siRNA transfection.

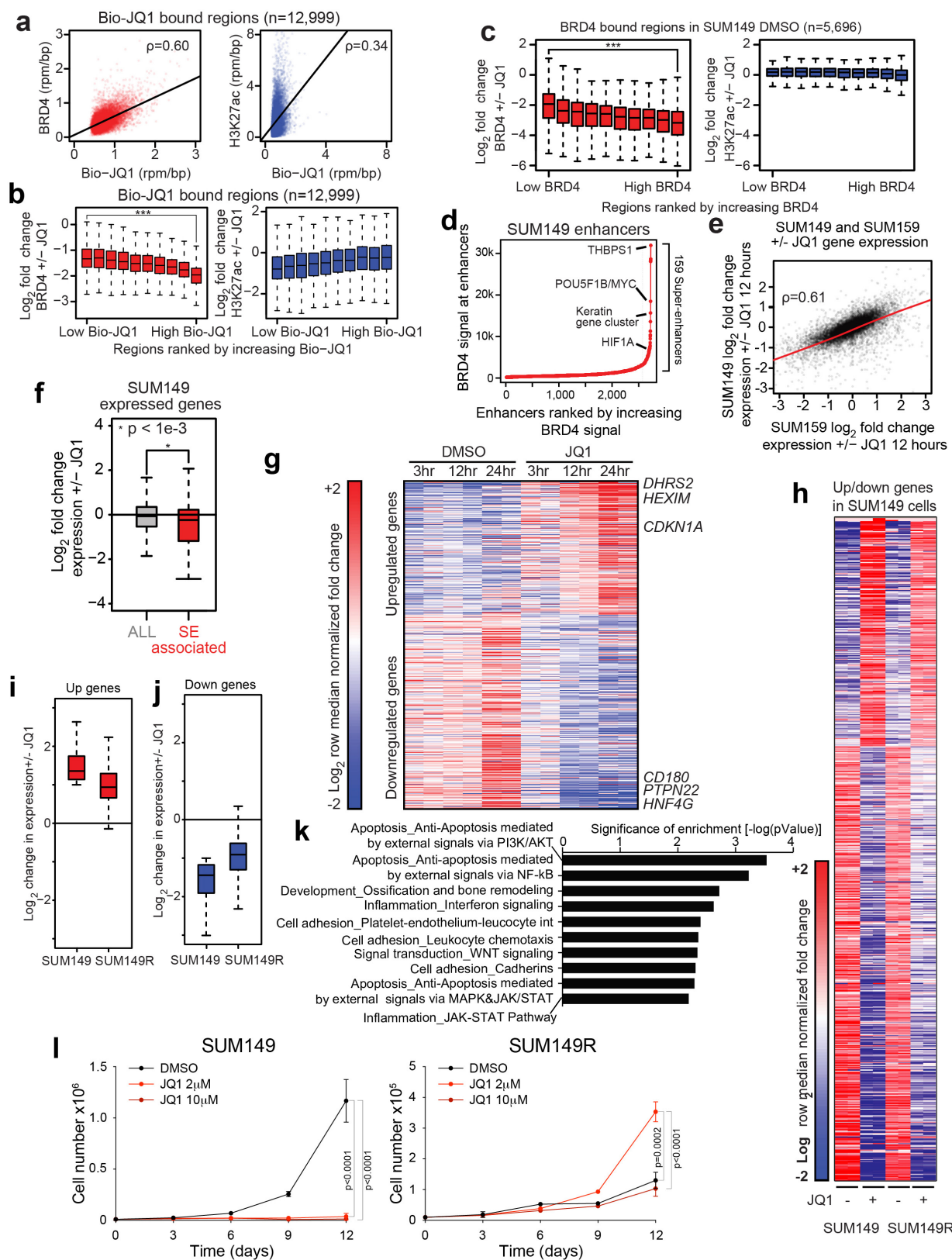
d, Cell cycle profile of SUM159 cells synchronized in G2/M with 100 ng ml^{-1} nocodazole followed by replating to fresh medium with DMSO or JQ1 (500 nM) added at -1 h or at 3 h after release. Cells were collected at different time points (0, 6, 12 h) after release. **e**, Immunoblot analysis of the indicated proteins at different time points (0, 3, 6, 12 h) after release of SUM159 cells synchronized with 100 ng ml^{-1} nocodazole followed by replating to fresh medium with DMSO or JQ1 (500 nM) added at 1 h before or 3 h after release. **f**, Cell cycle analysis of SUM159 cells following 72 h treatment with JQ1 (500 nM) or downregulation of BRD4 using TET-inducible shRNAs. **g**, Annexin V staining of SUM159 cells following 72 h treatment with JQ1 (500 nM) downregulation of BRD4 using TET-inducible shRNAs. All error bars represent s.e.m. **h**, Immunoblot analysis of the indicated proteins in a panel of breast cell lines; colour scheme as in **a**. For gel source data, see Supplementary Fig. 1.



Extended Data Figure 2 | See next page for figure caption.

Extended Data Figure 2 | Response to BBIs in TNBCs. **a**, Immunoblot analysis of the indicated proteins at different time points following JQ1 treatment (500 nM) in SUM159 cells (top) and at different JQ1 doses for 24 h treatment in SUM159 and MDA-MB-436 cells (bottom). **b**, Immunoblot analysis of the indicated proteins at different time points following JQ1 treatment (500 nM) in SUM149, SUM159 and MDA-MB-231 cells. **c**, Haematoxylin and eosin staining of SUM159 cells after 3 days of JQ1 treatment. **d**, Senescence β -galactosidase staining of SUM159 and MDA-MB-231 cells after 3 days of JQ1 treatment. Scale bars, 100 μ m. **e**, Box plots depict the weights of xenografts 30 days after injection of MDA-MB-231 (2×10^6) and IDC50X (2×10^5) cells into inguinal mammary fat pads of NOG mice; *n* indicates the number of mice per experiment. *P* values indicate statistical significance of the observed differences (unpaired *t*-test). Error bars represent s.e.m. Mice were administered JQ1 (50 mg per kg, daily) or vehicle only (control) for 14 days beginning at day 16 (MDA-MB-231) or 10 (IDC50X) after injection (after tumours reached palpable size). For EL12-58X PDX, mice were

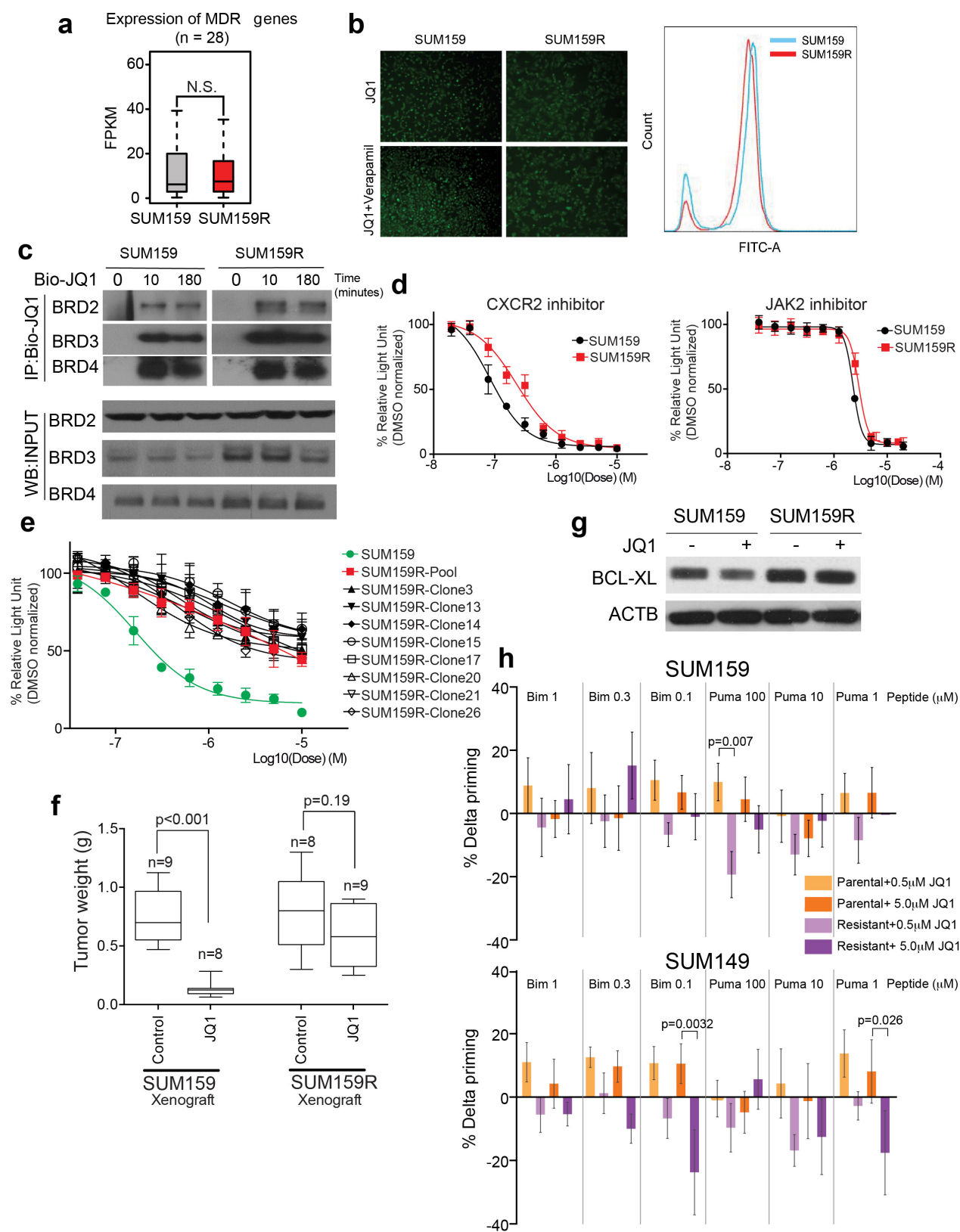
implanted with pieces of tissue measuring $1 \times 3 \times 3$ mm into the inguinal mammary fat pads and were administered daily JQ1 (50 mg per kg) for 14 days beginning at day 21 after injection (after tumours reached palpable size). **f**, Bromodeoxyuridine (BrdU) and luminal (Low MW CK) and basal high (High MW CK) molecular weight cytokeratin staining of EL12-58 xenograft with or without JQ1 treatment. Scale bars, 50 μ m. **g**, Tumour volume of SUM159 cells expressing TET-inducible BRD4-targeting shRNAs. Mice were administered doxycycline or vehicle only (control) for 39 days beginning at day 21 after injection (after tumours reached palpable size). Error bars represent s.d., *n* = 4 (shBRD4-1 experiment) and *n* = 5 (shBRD4-2 experiment). **h**, Haematoxylin and eosin staining and immunofluorescence analysis of basal (basal cytokeratin, cytokeratin 17, pSTAT3, and CD44) and luminal (luminal cytokeratin, cytokeratin 18, and CD24) markers in SUM159 xenografts with or without JQ1 treatment. Scale bars, 100 μ m for haematoxylin and eosin and 50 μ m for immunofluorescence, respectively. For gel source data, see Supplementary Fig. 1.



Extended Data Figure 3 | See next page for figure caption.

Extended Data Figure 3 | SUM149 JQ1 response. **a**, Scatter plots showing the relationship between the genomic binding of BRD4 and Bio-JQ1 (left) or H3K27ac and Bio-JQ1 (right) at all Bio-JQ1 enriched bound regions. Units of genomic occupancy are in rpm per bp. A simple linear regression is drawn in black. Pearson correlation statistics are also shown. **b**, Box plots showing the log₂ fold change in BRD4 with or without JQ1 (left) or H3K27ac with or without JQ1 (right) at Bio-JQ1 bound regions in SUM159 cells. The 12,999 Bio-JQ1 regions are ranked by increasing Bio-JQ1 binding and divided into 10 separate bins (displayed from left to right). The statistical significance of the difference in the mean BRD4 log₂ fold change between the weakest and strongest Bio-JQ1 bound region bins is shown (Welch's *t*-test; ****P* value < 1×10^{-10}). **c**, Box plots showing the log₂ fold change in BRD4 with or without JQ1 (left) or H3K27ac with or without JQ1 (right) at BRD4 bound regions in SUM149 cells. The 5,696 BRD4 bound regions are ranked by increasing background subtracted BRD4 binding and divided into 10 separate bins (displayed from left to right). The statistical significance of the difference in the mean BRD4 log₂ fold change between the weakest and strongest BRD4 bound region bins is shown (Welch's *t*-test; ****P* value < 1×10^{-10}). **d**, Ranked plots of enhancers defined in untreated SUM149 cells ranked by increasing BRD4 signal (units rpm). Enhancers are defined as regions of BRD4 binding not contained in promoters. The cut-off discriminating typical from super-enhancers is shown as a dashed grey line. Enhancers associated with TNBC characteristic genes are highlighted. **e**, Scatter plots showing the relationship between the log₂ fold change in gene expression upon 12 h JQ1 treatment in SUM149 (*y*-axis) and SUM159 (*x*-axis) cells. A simple linear regression is drawn in red. The Pearson

correlation statistic is also shown. **f**, Box plots showing the log₂ fold change in expression relative to DMSO control of either all active genes or super-enhancer (SE)-associated genes upon 12 h JQ1 treatment. The statistical significance of the difference in expression change between all active genes and super-enhancer-associated genes is shown by a Welch's *t*-test **P* value < 1×10^{-3}). **g**, Heat map showing the expression of genes that are up or down regulated by JQ1 versus DMSO after 24 h treatment. Each row shows the expression of a single gene in either DMSO or JQ1 treated cells at 3, 12, and 24 h after treatment. Expression values are coloured according to fold change relative to the median for each row. Genes are ordered by fold change with or without JQ1 24 h after treatment. **h**, Heat map showing the expression of genes that are up or down regulated by JQ1 versus DMSO after 12 h treatment in SUM149 and SUM149R cells. Each row shows the expression of a single gene in either DMSO or JQ1 treated cells at 12 h after treatment. Expression values are coloured according to fold change relative to the median for each row. Genes are ordered by fold change with or without JQ1 12 h after treatment in SUM149 cells. **i**, **j**, Box plots showing the log₂ fold change in expression at genes that are up- (**i**) or down- (**j**) regulated by JQ1 versus DMSO after 12 h of treatment in parental SUM149 cells. Log₂ fold change in expression is shown for either parental SUM149 (left) or resistant SUM149R (right) cells. **k**, Top signalling pathways affected by JQ1-induced gene expression changes in SUM159 cells. **l**, Viable cell numbers of SUM149 (left) and SUM149R (right) cells treated with different doses of JQ1 (2 μM, 10 μM). Error bars represent s.d., *n* = 3. *P* values indicate statistical significance of the observed differences (two-way ANOVA with Bonferroni multiple comparison correction).

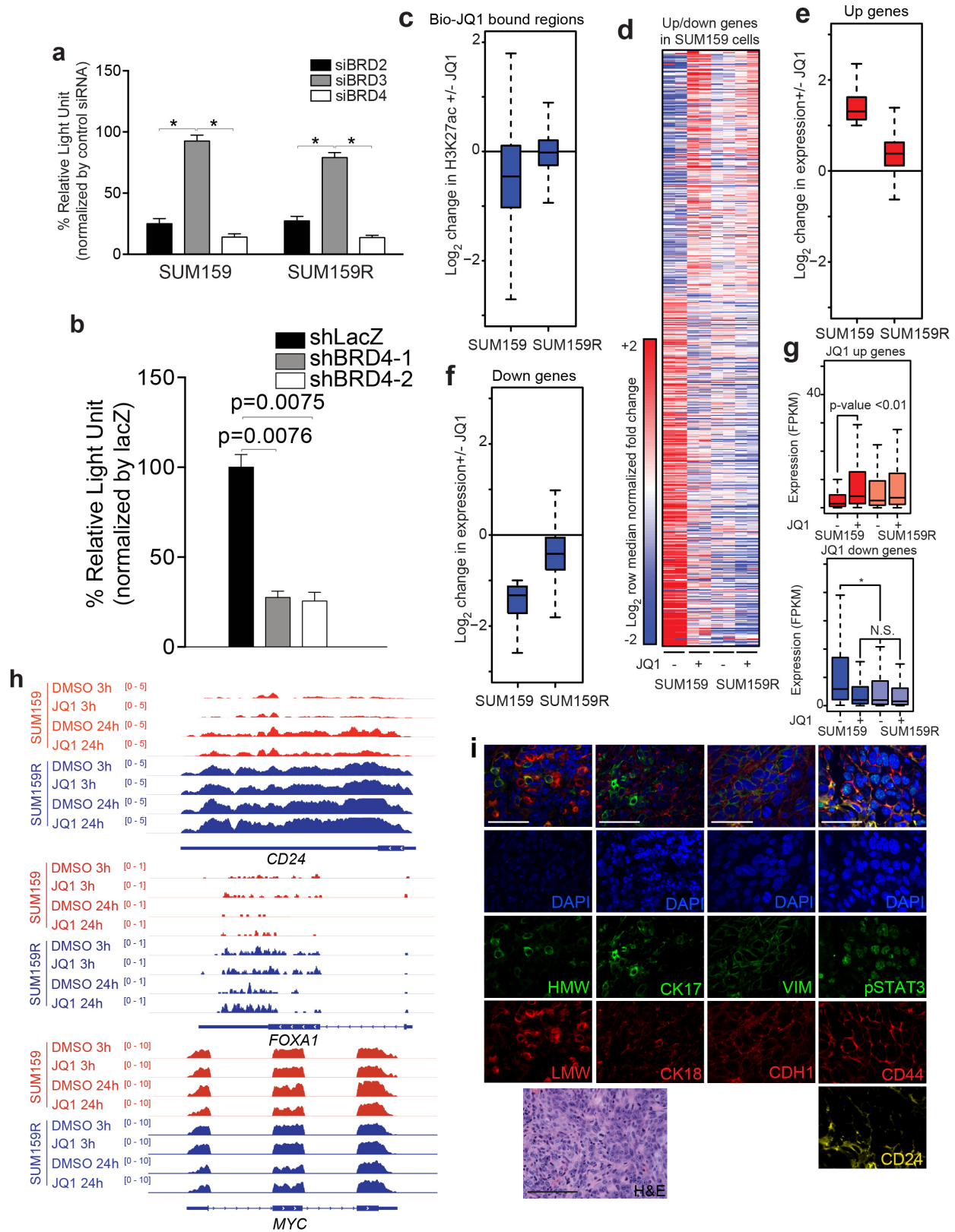


Extended Data Figure 4 | See next page for figure caption.

Extended Data Figure 4 | Characterization of SUM159R cells.

a, Expression of ABC transporters in SUM159 and SUM159R cells. The expression of 29 ABC transporters was analysed based on RNA-seq data on the two cell lines. **b**, Assay for MDR (multi drug resistance) pumps in SUM159 and SUM159R cells treated with JQ1 alone or together with verapamil based on microscopic examination (left) and FACS (right) of cells labelled with fluorescent MDR substrate. **c**, Immunoprecipitation analysis of Biotinylated JQ1 (Bio-JQ1) in SUM159 and SUM159R cells with JQ1 treatment at different time points following immunoblot for the indicated proteins. **d**, Cellular viability of SUM159 and SUM159R cells treated with CXCR2 and JAK2 inhibitors. Error bars represent s.d., $n = 3$. **e**, Cellular viability of SUM159, and pool and single cell clones of SUM159R cells treated with different doses of JQ1. Error bars represent s.d., $n = 3$. **f**, Tumour weight of xenografts derived from SUM159 and

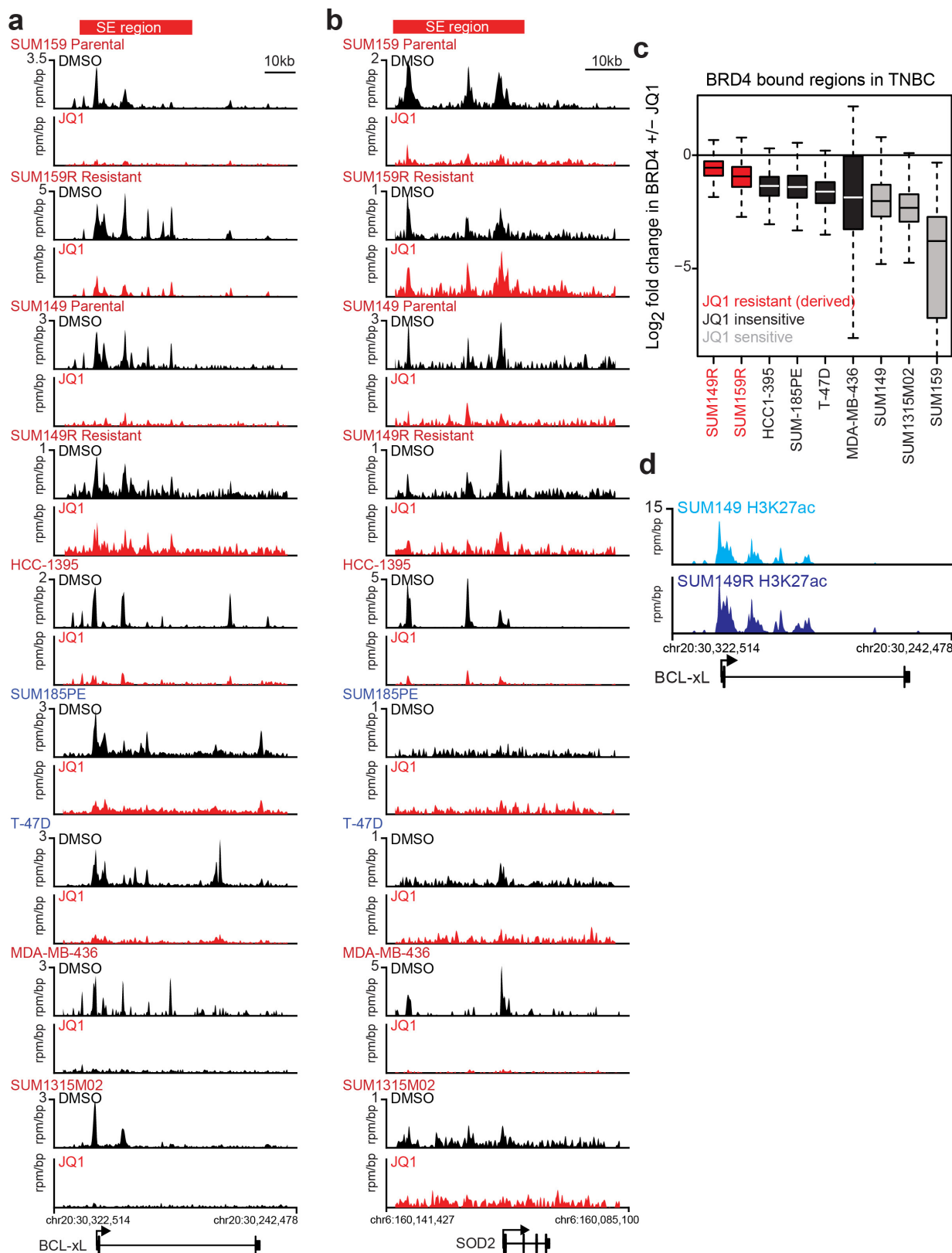
SUM159R cells. Mice were administered JQ1 for 14 (SUM159) and 30 (SUM159R) days beginning at day 14 and 26, respectively, after injection. *P* values indicate statistical significance of the observed differences (unpaired *t*-test). Error bars represent s.e.m. **g**, Immunoblot analysis of BCL-XL expression in SUM159 and SUM159R cells before and after JQ1 3 h treatment (500 nM). **h**, Dynamic BH3 profiling reveals inverse apoptotic response to JQ1 in SUM149R and SUM159R cells. In parental lines JQ1 increases priming relative to untreated cells indicating an increase in apoptotic propensity. In resistant lines JQ1 reduces priming indicating greater resistance to apoptosis relative to untreated cells. *P* values indicate statistical significance of the observed differences (two-way ANOVA). Error bars represent s.e.m., $n = 5$. For gel source data, see Supplementary Fig. 1.



Extended Data Figure 5 | See next page for figure caption.

Extended Data Figure 5 | BRD4 binding in SUM159R cells. **a**, Cellular viability of SUM159 and SUM159R cells transfected with siRNAs targeting bromodomain proteins. Asterisks indicate statistical significance (paired *t*-test) of the marked differences as follows: SUM159: siBRD2 versus siBRD3, $P = 0.013$, siBRD3 versus siBRD4, $P = 0.0154$ and SUM159R: siBRD2 versus siBRD3, $P = 0.0159$, siBRD2 versus siBRD4, $P = 0.0048$; siBRD3 versus siBRD4, $P = 0.0068$. **b**, Cellular viability of SUM159R cells expressing TET-inducible BRD4-targeting or lacZ shRNAs. All error bars represent s.e.m. *P* values indicate statistical significance of the observed differences (unpaired *t*-test). **c**, Box plot showing the log₂ fold change in H3K27ac genomic occupancy at regions bound by Bio-JQ1 in parental SUM159 or resistant SUM159R cells. **d**, Heat map showing the expression of genes that are up or down regulated by JQ1 versus DMSO after 24 h treatment in parental SUM159 cells. Each row shows the expression of a single gene in either DMSO or JQ1 treated cells at 24 h after treatment in SUM159 cells (left four columns) or SUM159R cells (right four columns). Expression values are coloured according to fold change relative to the median for each row. Genes are ordered by fold change with or without JQ1 24 h after treatment. **e**, **f**, Box plots showing the log₂ fold change in expression at genes that are up- (**e**) or down- (**f**) regulated by JQ1

versus DMSO after 24 h of treatment in parental SUM159 cells. Log₂ fold change in expression is shown for either parental SUM159 or resistant SUM159R cells. **g**, Box plots showing expression of genes that are up- or downregulated by JQ1 versus DMSO after 24 h of treatment in parental SUM159 cells. Expression is shown in DMSO and JQ1-treated conditions in units of FPKM for either parental SUM159 (left) or resistant SUM159R (right) cells. The statistical significance of the difference between gene expression distributions for SUM159 DMSO- and JQ1-treated cells is shown ($P < 0.01$). The difference between all other distributions are considered non significant (NS). The statistical significance of the difference between SUM159 DMSO gene expression distribution and all other distributions is shown ($*P < 1 \times 10^{-3}$). The difference between all other distributions are considered non-significant. **h**, Examples of luminal and basal cell-specific genes, and MYC in SUM159 and SUM159R cells. RNA-seq tracks are shown. **i**, Haematoxylin and eosin staining and immunofluorescence analysis of luminal (CK18 and LMW) and basal (CK17 and HMW) cytokeratins and luminal (VIM and CD24) and basal (CDH1, CD44, and pSTAT3) cell markers in SUM159R xenografts. All error bars represent s.e.m. Scale bars, 100 μ m for haematoxylin and eosin and 50 μ m for immunofluorescence, respectively.

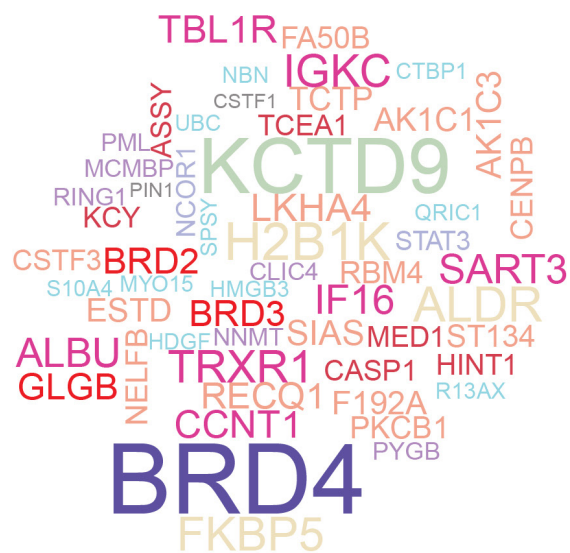


Extended Data Figure 6 | JQ1 response in other breast cancer cell lines.

a, b, Gene tracks depicting BRD4 + DMSO and BRD4 + JQ1 in multiple breast cancer cells at the *BCL-xL* (**a**) or *SOD2* (**b**) gene loci. The x-axis shows position along the chromosome with gene structures drawn below. The y-axis shows genomic occupancy in units of rpm per bp. The *BCL-xL* and *SOD2* super-enhancers are shown as a red bar at the top. **c,** Box plots showing the log₂ fold change in BRD4 occupancy with or without JQ1 for

all BRD4 bound regions in each cell line for multiple TNBC. Cell lines are ordered by their JQ1 (IC₅₀) and coloured by their sensitivity. **d,** Gene tracks depicting H3K27AC occupancy at the *BCL-xL* locus in SUM149 parental (top, light blue) or SUM149R resistant (bottom, dark blue) cells. The x-axis shows position along the chromosome with gene structures drawn below. The y-axis shows genomic occupancy in units of rpm per bp. All error bars represent s.e.m.

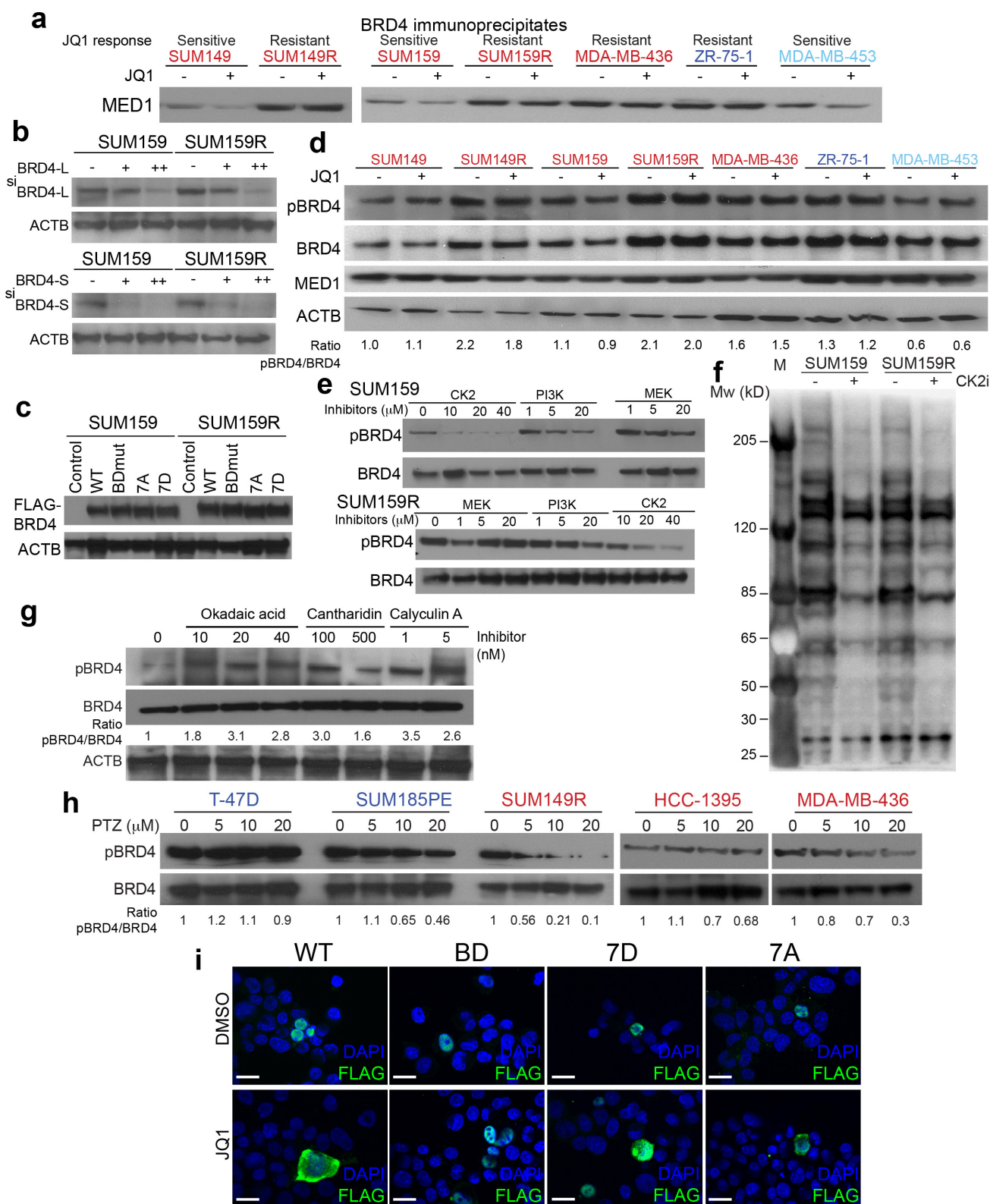
SUM159



SUM159R

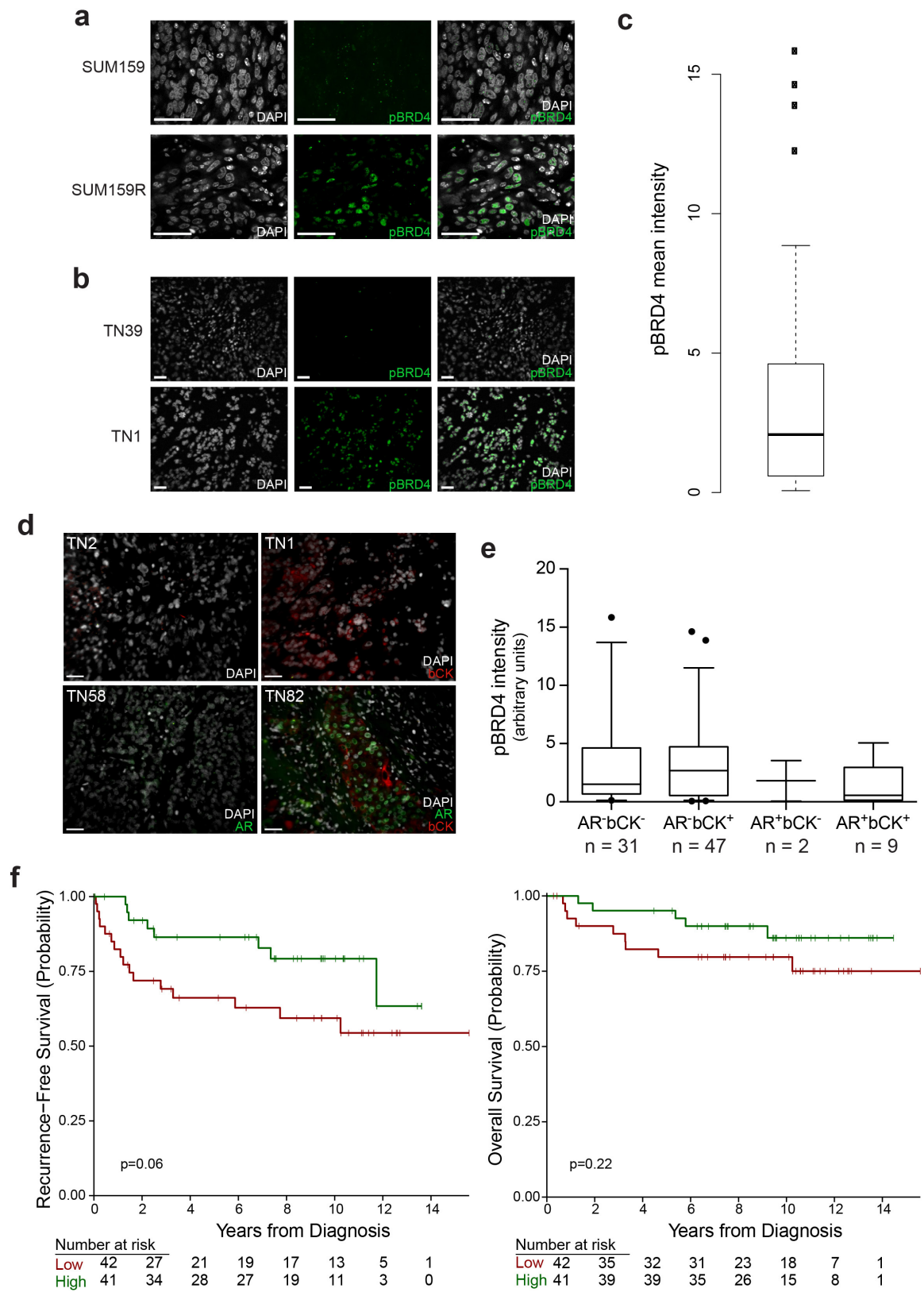


Extended Data Figure 7 | Word clouds depicting BRD4-associated proteins identified in RIME analysis.



Extended Data Figure 8 | Mechanism of BBI-resistance. **a**, Immunoblot analysis of BRD4 immunoprecipitates for MED1 in the indicated cell lines with or without JQ1 treatment (5 μM, 3 h). **b**, Immunoblot analysis of long (BRD4L) and short (BRD4S) forms of BRD4 after transfection of siRNAs. **c**, Immunoblot analysis of the indicated exogenously expressed Flag-tagged BRD4 proteins in SUM159 and SUM159R cells. **d**, Immunoblot analysis of pBRD4, BRD4, MED1 and ACTB in the indicated cell lines with or without JQ1 treatment. **e**, Immunoblot analysis of phospho-BRD4 (pBRD4) and BRD4 in SUM159 and SUM159R cells treated with the indicated doses of CK2, PI3K, and MEK inhibitors for 2 h.

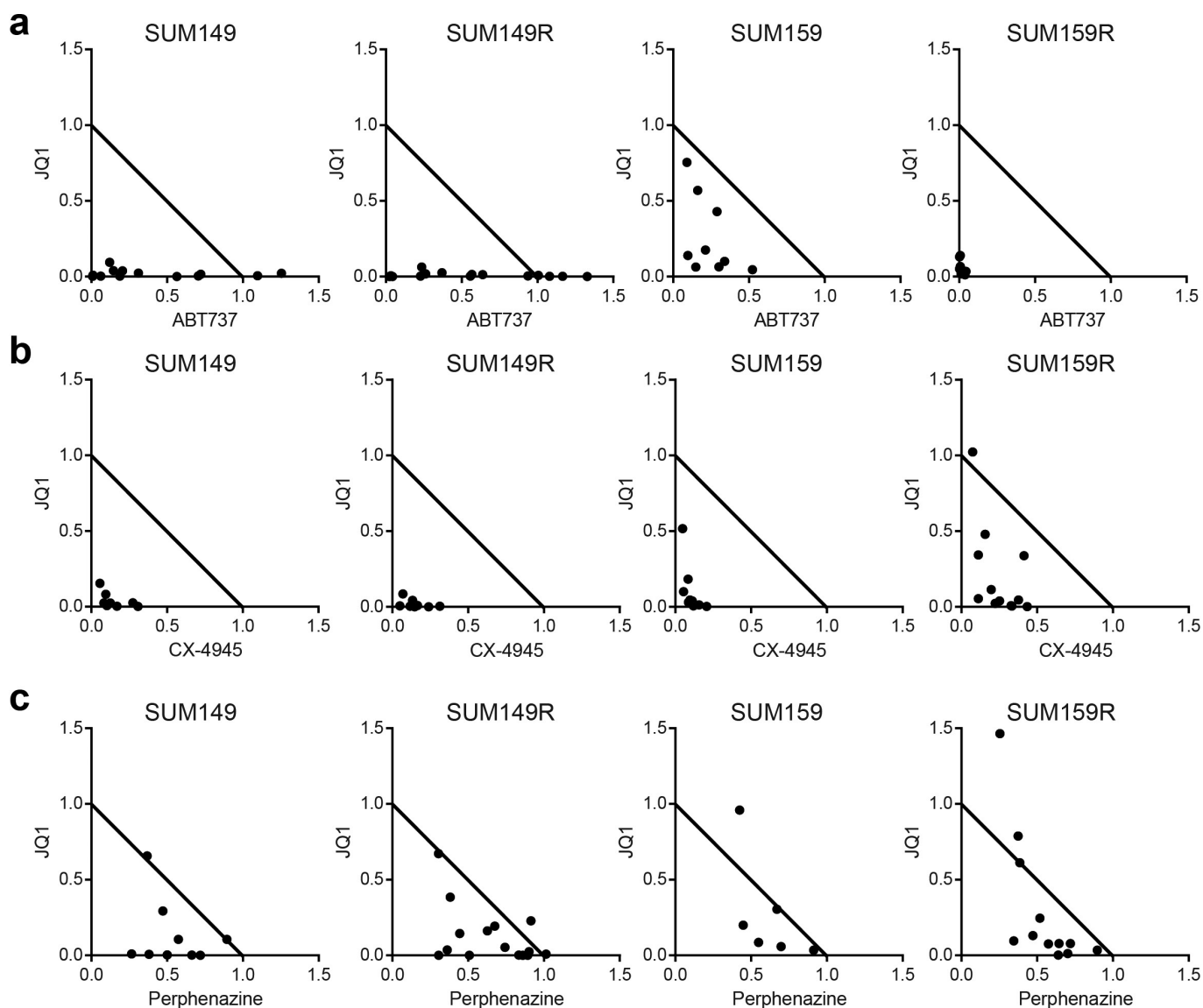
f, Immunoblot analysis of CK2 substrates in SUM159 and SUM159R cells following CK2 inhibitor (CX-4945, 10 μM) 3 h treatment. **g**, Immunoblot analysis of pBRD4 and BRD4 in SUM149 cell line treated with different doses of the indicated PP2A inhibitors for 3 h. ACTB was used as loading control. **h**, Immunoblot analysis of pBRD4 and BRD4 in the indicated cell lines treated with different doses of phenothiazine for 6 h. **i**, Immunofluorescence analysis of exogenous Flag-tagged BRD4 proteins (WT, BD, 7D and 7A) in SUM159 cells with or without JQ1 treatment (5 μM, 3 h). Scale bars, 20 μm. For gel source data, see Supplementary Fig. 1.



Extended Data Figure 9 | See next page for figure caption.

Extended Data Figure 9 | Phospho-BRD4 levels in xenografts and primary TNBC samples. **a**, Immunofluorescence analysis of phospho-BRD4 (pBRD4) in SUM159 parental and SUM159R xenografts showing that resistance is associated with higher pBRD4 levels. **b**, Examples of pBRD4 immunofluorescence in patient tumours depicting variability among different TNBC samples. Scale bars, 50 μ m. **c**, Mean intensity of phospho-BRD4 (pBRD4) in tissue samples from 83 patients with early-stage triple negative breast cancer (TNBC). **d**, Examples of androgen receptor and basal cytokeratin (bCK, HMW CK) immunofluorescence in TNBC samples. Scale bars, 50 μ m. **e**, Box plot depicting pBRD4 signal intensity in TNBCs tumours with the indicated androgen receptor and bCK expression patterns. None of the differences among groups were

significant (ANOVA test $P = 0.5413$ and Dunnett's multiple comparisons test was not significant). **f**, Kaplan–Meier estimates of disease-free survival (DFS) and overall survival (OS) in TNBC subgroups using a median-split of pBRD4 intensity. Disease outcomes were evaluated in 83 of 89 TMA samples. Patients with low pBRD4 had a worse overall prognosis with a five-year RFS of 66.2% (95% confidence interval (CI) 52.7–83.1%), compared to an RFS of 86.4% (95% CI 76.0–98.3%) among patients with high pBRD4 (HR = 2.3, 95% CI 0.98–5.4, $P = 0.06$). However, with this small sample size this difference did not reach statistical significance, nor did the association of pBRD4 status and overall survival (HR = 2.0, 95% CI 0.67–5.9, $P = 0.22$).



Extended Data Figure 10 | Overcoming BBI-resistance. a–c, Synergy studies of JQ1 with ABT737 (BCL-x1 and BCL-2 inhibitor) (a), CX-4945 (CK2 inhibitor) (b) and perphenazine (PP2A activator) (c). Points represent paired values of drug concentrations assessed for synergism.

The diagonal line signifies drug additivity. Points above the line represent antagonistic drug combinations, and those below the line represent synergistic drug combinations.

Super-resolution imaging reveals distinct chromatin folding for different epigenetic states

Alistair N. Boettiger¹, Bogdan Bintu¹, Jeffrey R. Moffitt¹, Siyuan Wang¹, Brian J. Beliveau², Geoffrey Fudenberg³, Maxim Imakaev³, Leonid A. Mirny³, Chao-ting Wu² & Xiaowei Zhuang¹

Metazoan genomes are spatially organized at multiple scales, from packaging of DNA around individual nucleosomes to segregation of whole chromosomes into distinct territories^{1–5}. At the intermediate scale of kilobases to megabases, which encompasses the sizes of genes, gene clusters and regulatory domains, the three-dimensional (3D) organization of DNA is implicated in multiple gene regulatory mechanisms^{2–4,6–8}, but understanding this organization remains a challenge. At this scale, the genome is partitioned into domains of different epigenetic states that are essential for regulating gene expression^{9–11}. Here we investigate the 3D organization of chromatin in different epigenetic states using super-resolution imaging. We classified genomic domains in *Drosophila* cells into transcriptionally active, inactive or Polycomb-repressed states, and observed distinct chromatin organizations for each state. All three types of chromatin domains exhibit power-law scaling between their physical sizes in 3D and their domain lengths, but each type has a distinct scaling exponent. Polycomb-repressed domains show the densest packing and most intriguing chromatin folding behaviour, in which chromatin packing density increases with domain length. Distinct from the self-similar organization displayed by transcriptionally active and inactive chromatin, the Polycomb-repressed domains are characterized by a high degree of chromatin intermixing within the domain. Moreover, compared to inactive domains, Polycomb-repressed domains spatially exclude neighbouring active chromatin to a much stronger degree. Computational modelling and knockdown experiments suggest that reversible chromatin interactions mediated by Polycomb-group proteins play an important role in these unique packaging properties of the repressed chromatin. Taken together, our super-resolution images reveal distinct chromatin packaging for different epigenetic states at the kilobase-to-megabase scale, a length scale that is directly relevant to gene regulation.

Multiple lines of evidence suggest that the spatial organization of chromatin at the kilobase-to-megabase scale is important for genomic functions^{2–4,6–8,10–12}. The sizes of genes, gene clusters and regulatory domains all occur in this range; in addition, physical interactions between genomic elements separated by this distance range, such as promoter–enhancer interactions, are important for gene activity^{2–4}. Recent high-throughput chromatin conformation capture measurements revealed that individual chromosomes are partitioned into contact domains or topologically associating domains with lengths ranging from tens of kilobases (kb) to multiple megabases (Mb), and that this structural organization could be relevant to a variety of genome functions^{3,6–8,10–12}. Chromatin is also demarcated into domains of distinct epigenetic states characterized by biochemical modifications and DNA-binding proteins^{9–11}. Yet, how the 3D spatial organization of chromatin differs amongst these different epigenetic states is largely unknown.

Direct imaging of the spatial organization of chromatin in different epigenetic states requires the ability to specifically label genomic DNA *in situ* and to image chromatin structures with a high resolution. Here we used fluorescence *in situ* hybridization (FISH) to label specific regions of the genome with complementary oligonucleotide probes tagged with fluorescent dyes. We adopted and modified a previously reported Oligopaint approach^{13,14} to produce tens of thousands of unique, oligonucleotide probes to label kilobase-to-megabase long genomic regions using massively parallel oligonucleotide synthesis^{13–16}. High-yield probe synthesis was achieved with a recently described enzymatic amplification method¹⁷ (Extended Data Fig. 1 and Supplementary Methods). We used osmotically balanced fixation conditions to minimize shrinkage effect (Supplementary Methods) and observed no detectable chromatin shrinkage (Extended Data Fig. 2). We then imaged the labelled chromatin regions using a super-resolution imaging method, three-dimensional stochastic optical reconstruction microscopy (3D-STORM)^{18,19}. This approach yielded images of specific genomic regions in cells with 20-nm *xy* and 50-nm *z* resolution¹⁴.

We imaged 46 epigenetically defined genomic domains (Extended Data Fig. 3a; Extended Data Table 1) in *Drosophila* Kc167 cells. We classified these regions into three major epigenetic states—transcriptionally active, inactive, and Polycomb-repressed (hereafter referred to as Active, Inactive and Repressed, respectively)—on the basis of enrichment of histone modifications and regulatory proteins from ChIP-seq (chromatin immunoprecipitation followed by sequencing) and DamID (DNA adenine methyltransferase identification) data (Fig. 1a), as described previously^{10,11,20} and in Supplementary Methods. Active chromatin domains were selected based on the enrichment of the histone modifications H3K4me2 or H3K79me3. Repressed domains were selected based on enrichment for H3K27me3 or Polycomb Group (PcG) proteins. Inactive domains were selected based on the predominance of unmodified histones and a depletion of PcG proteins and transcriptional activators. The lengths of the selected domains span those observed for all three epigenetic types in *Drosophila* (~10–500 kb).

Compared to conventional fluorescence images, STORM images of these domains revealed substantially more structural information (Fig. 1b and Extended Data Fig. 4). From these super-resolution images, we first measured the physical volume occupied by each domain, as a measure of the compactness of the DNA in the domain (Fig. 1c; Extended Data Fig. 3b, c). The volume measurements showed cell-to-cell variations (Extended Data Fig. 5a, b), which may reflect differences in expression state or cell cycle. However, the values averaged over ~50 randomly selected cells for each domain exhibited a clear difference among the three types of epigenetic domains. Across the nearly two orders of magnitude of domain lengths that we studied, the median volumes of the Active domains were always larger than those of the Inactive domains of the same domain lengths, which were

¹Howard Hughes Medical Institute, Department of Chemistry and Chemical Biology, Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA. ²Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ³Institute for Medical Engineering and Science, and Department of Physics, Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts 02139, USA.

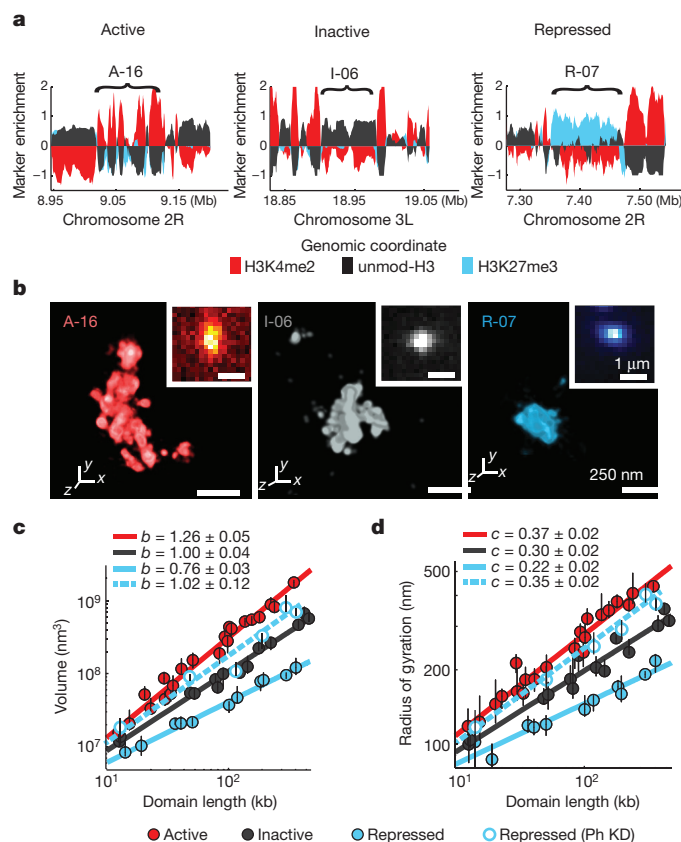


Figure 1 | Chromatin in different epigenetic states exhibits distinct packaging and power-law scaling. **a**, Enrichment profile of H3K4me2 (red), H3K27me3 (light blue) and unmodified H3 (black) in three genomic regions, each harbouring an example Active, Inactive or Repressed domain (indicated by brackets). Marker enrichment, as defined in Supplementary Methods, was determined from ChIP-seq data²⁰. **b**, 3D-STORM images of the three distinct epigenetic domains in **a**, labelled by *in situ* hybridization with DNA probes conjugated to the photoswitchable dye Alexa-647, shown with their corresponding conventional images in the inset. Each epigenetic domain appears as a single region in nearly all cells due to homologous pairing in the tetraploid Kc167 cells. **c**, log-log plot of the median domain volume as a function of domain length for Active (red solid circles), Inactive (black solid circles) and Repressed (light blue solid circles) domains, as well as for Repressed domains in Ph-knockdown (Ph KD) cells (light blue hollow circles). Error bars represent 95% confidence intervals derived from resampling ($n \approx 50$ cells). The lines indicate power-law fits, with the scaling exponent b shown in the legend. **d**, As in **c** but for the radius of gyration as a function of domain length with the scaling exponent c shown in the legend.

in turn always larger than those of the Repressed domains (Fig. 1c, solid circles). These results are in line with previous data showing that PcG proteins can lead to chromatin compaction^{2,21–23} and that actively transcribed chromatin regions tend to be more open than non-transcribing regions².

Notably, the volume (V) of the chromatin domains exhibited a power-law scaling behaviour with the domain length (L), that is, $V \propto L^b$, and the scaling exponent b was distinct for the three different epigenetic states (Fig. 1c, solid circles; Extended Data Fig. 3b, c). Inactive chromatin domains had a scaling exponent of $b = 1.00 \pm 0.04$ (\pm standard error), indicating that the 3D density of the chromatin was constant over different domain lengths. Active domains had a scaling exponent significantly greater than 1 ($b = 1.26 \pm 0.05$), indicating increasingly less dense packaging for larger domains. Repressed domains exhibited a scaling exponent that was notably less than 1 ($b = 0.76 \pm 0.03$), indicating that the packaging density increased with increasing domain length. As an alternative measure of the physical sizes of chromatin domains, we determined the radius of gyration (R_g), defined as the

root-mean square distance of molecule positions measured by STORM in each domain from the centroid of these positions in the domain (Supplementary Methods). Power-law scaling was also observed for R_g as a function of L , that is, $R_g \propto L^c$, with the scaling exponents $c = 0.37 \pm 0.02$, 0.30 ± 0.02 , and 0.22 ± 0.02 , for Active, Inactive and Repressed domains, respectively (Fig. 1d, solid circles; Extended Data Fig. 3d). These scaling behaviours were conserved across different genomic regions on multiple chromosomes (Extended Data Fig. 3a, b), suggesting that the different packaging behaviours are characteristic of the epigenetic states. Epigenetic states also influence the scaling of contact frequencies measured by chromosome conformation capture²⁴, but how contact frequency is related to the size measurements here remains to be understood. In addition to different size-scaling properties, we found that these different types of epigenetic domains also tend to have different 3D shape characteristics (Extended Data Fig. 3e, f).

Next we probed how chromatin was folded within epigenetic domains. To this end, we selected two large chromatin domains for each epigenetic type and measured the R_g of internal regions of varying lengths within these domains, hereafter referred to as subdomains (Fig. 2a, b; Extended Data Fig. 5c; Extended Data Table 1). Interestingly, both Inactive and Active domains showed a self-similar organization, in which the internal subdomains exhibited scaling behaviours that were similar to those observed for the whole epigenetic domains (Fig. 2b, left and middle). In stark contrast, we did not observe such a self-similar organization for either of the Repressed chromatin domains investigated (the Bithorax (Fig. 2b, right) and Antennapedia (Extended Data Fig. 6) complexes). Instead, the R_g values grew rapidly as a function of subdomain length and quickly saturated, such that subdomains longer than approximately one fifth of the length of the parent domain essentially all exhibited the same R_g values.

The observation that even a small subdomain traversed nearly the entire volume of the parent domain predicts that two such small subdomains would occupy the same physical space, suggesting a high degree of intermixing of chromatin within these Repressed domains. We tested this hypothesis by simultaneously labelling two subdomains within the same Repressed domain with two distinct sets of FISH probes conjugated to spectrally distinct photoswitchable dyes and imaged these subdomains with two-colour STORM (Fig. 2c, right panel). Indeed, images of these subdomain pairs showed a high degree of overlap, markedly distinct from the behaviours observed for the Active and Inactive chromatin regions (Fig. 2c; left and middle panels). Quantitatively, the pairs of subdomains within each Repressed domain showed ~ 60 – 80% overlap in space (median 68%, 3 different pairs of subdomains investigated, $n \approx 150$ cells) (Fig. 2d, light blue). In contrast, the neighbouring subdomains of Inactive chromatin only showed approximately 10–30% spatial overlap (median 26%, 3 different pairs of subdomains, $n \approx 150$ cells) (Fig. 2d, black) and the neighbouring subdomains of Active chromatin only showed approximately 15–25% spatial overlap (median 20%, 2 different pairs of subdomains, $n \approx 100$ cells) (Fig. 2d, red). The difference observed between Active (or Inactive) subdomains and Repressed subdomains is statistically highly significant ($P < 1 \times 10^{-10}$, Wilcoxon test). These results indicate that the degree of intermixing of chromatin within individual epigenetic domains depends strongly on the epigenetic state.

We then probed how these different epigenetic domains interacted with one another across epigenetic boundaries. Notably, the Repressed domains did not show any appreciable overlap with neighbouring Active domains, whereas the neighbouring Inactive and Active domains partially intermixed with each other (Fig. 3a, b). We quantified four different Repressed–Active boundaries and three different Inactive–Active boundaries. The Repressed domains typically showed less than 3% overlap with their neighbouring Active domains (median 1.5%, $n \approx 150$ cells), whereas Inactive domains exhibited up to 15% overlap with neighbouring Active domains (median 9.8%, $n \approx 150$) (Fig. 3c). The difference between these two types of domain boundaries was statistically highly significant ($P < 1 \times 10^{-14}$, Wilcoxon test). Therefore,

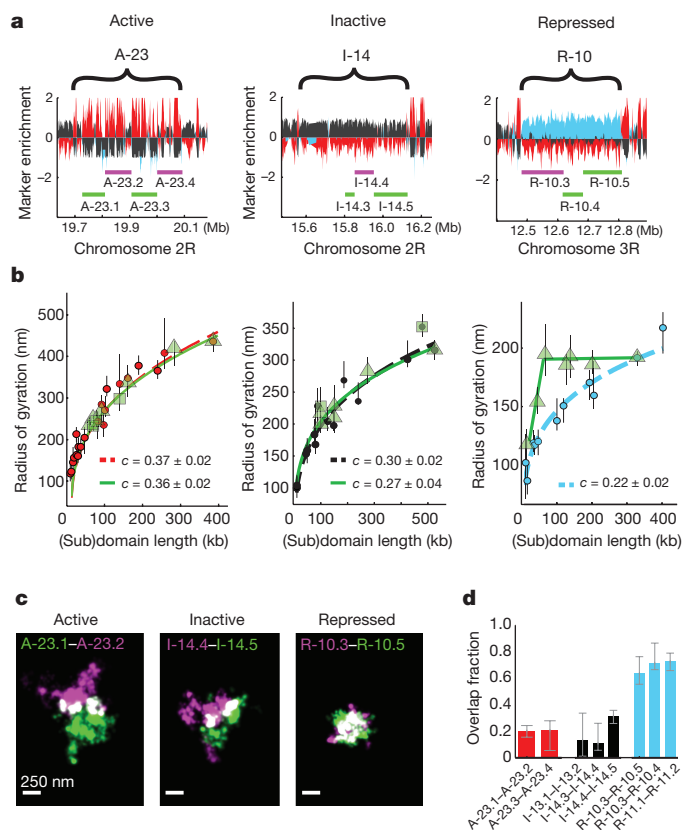


Figure 2 | Different types of epigenetic domains exhibit distinct subdomain scaling and intermixing. **a**, Marker enrichment profile of three genomic regions with the example epigenetic domains marked by brackets and imaged subdomains marked by green and magenta lines. **b**, Linear plot of the radius of gyration as a function of the subdomain length (green symbols), compared to those for the whole domain data (red, black or light blue circles), for Active (left), Inactive (middle) and Repressed chromatin (right). Different green symbols (triangle and squares) represent subdomains of two different parent domains. Power-law fits of subdomains (green solid lines) and whole domains (red, black and light blue dashed lines) are shown with the scaling exponent c given in the legends. The green lines in the right panel are to guide the eye. **c**, Two-colour, 3D-STORM images of example pairs of subdomains within Active (left), Inactive (middle), and Repressed (right) domains. Portions of the two subdomains that overlap in 3D are shown in white. The two subdomains are labelled with Alexa-647- and Alexa-750-tagged DNA probes, respectively. **d**, Quantification of overlap fraction between the subdomains for Active (red), Inactive (black), or Repressed (light blue) chromatin (Supplementary Methods). Error bars represent 95% confidence intervals derived from resampling ($n \approx 50$ cells).

the degree of spatial separation between neighbouring domains of different epigenetic types also depends strongly on the epigenetic state.

The different packaging and intermixing behaviours observed for these chromatin types point to distinct mechanisms involved in chromatin folding in the different epigenetic states. It is remarkable that the two types of non-transcribing chromatin—Inactive and Repressed states—exhibited such distinct packaging behaviours. The Repressed chromatin appeared to exhibit a substantially more compact packing, higher degree of chromatin intermixing within domains, and stronger tendency to spatially exclude neighbouring domains. To explore the potential mechanisms underlying these different packaging behaviours, we employed stochastic polymer simulations. It has been suggested that genomic DNA could be approximated as a fractal globule, which arises when an unknotted polymer is confined into a volume substantially smaller than the relaxed volume of the polymer^{25–27}. Our experimentally observed power-law scaling ($R_g \propto L^{0.3}$) and self-similar organization of the Inactive chromatin are roughly consistent with the expected

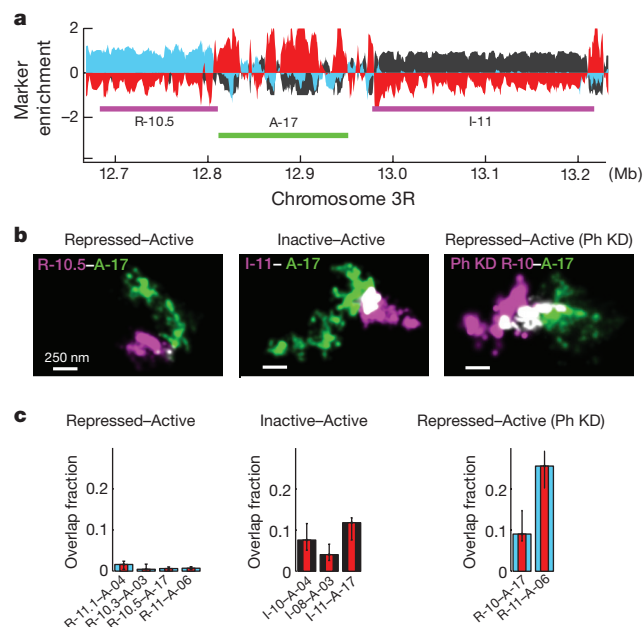


Figure 3 | Neighbouring chromatin domains show different amount of intermixing for different types of epigenetic boundaries. **a**, Marker enrichment profile of a genomic region harbouring three epigenetic domains marked by the magenta and green lines. **b**, Two-colour, 3D-STORM images for neighbouring epigenetic domains with Repressed-Active and Inactive-Active boundaries in wild-type cells, as well as for the Repressed-Active boundaries in Ph-knockdown (Ph KD) cells. **c**, Quantification of overlap fraction between the indicated neighbouring domains (Supplementary Methods). Error bars represent 95% confidence intervals derived from resampling ($n \approx 50$ cells).

properties of the fractal globule, as confirmed by our simulations of a polymer confined to a small volume (Supplementary Methods) (Fig. 4a, b).

Polycomb-repressed chromatin, on the other hand, exhibited distinct folding behaviours that could not be explained by the fractal globule model. Prompted by the observations that some PcG proteins can bridge nucleosomes^{21,22}, we used self-interacting monomers²⁸ to simulate such PcG-mediated chromatin interactions. We then embedded such a ‘sticky’ polymer domain containing self-interacting monomers within a large non-sticky polymer (to emulate the surrounding non-repressed domains), and simulated the polymer in a confined volume (Supplementary Methods). This model reproduced the packaging behaviours that we observed for Repressed chromatin domains provided that the monomer-interaction strength was not too strong to cause irreversible monomer binding and that the simulation was sufficiently long to achieve intermixing (Supplementary Methods). First, the scaling exponent of 0.22 ± 0.01 ($R_g \propto L^{0.22}$) derived from the simulation (Fig. 4c) agreed quantitatively with the experimentally observed value (Fig. 1d, solid blue line). Second, subdomains of these simulated polymers showed a saturation behaviour where the subdomains had nearly the same R_g values as the full domain, until they became less than one third the length of the parent domain (Fig. 4d), also similar to our experimental observation (Fig. 2b, right panel). Third, subdomains within the sticky polymer domains exhibited substantially more intermixing than neighbouring subdomains of the non-sticky polymer, and the sticky polymer domains spatially excluded neighbouring non-sticky regions (Fig. 4e, f), both behaviours also consistent with the differences that we observed experimentally between Repressed and Inactive chromatin domains (Figs 2c, d and 3b, c). Overall, our simulation results suggest that interactions between PcG proteins could provide an explanation for the distinct folding behaviour observed for the Repressed chromatin, although it is possible that other mechanisms also contribute to the observed folding behaviour.

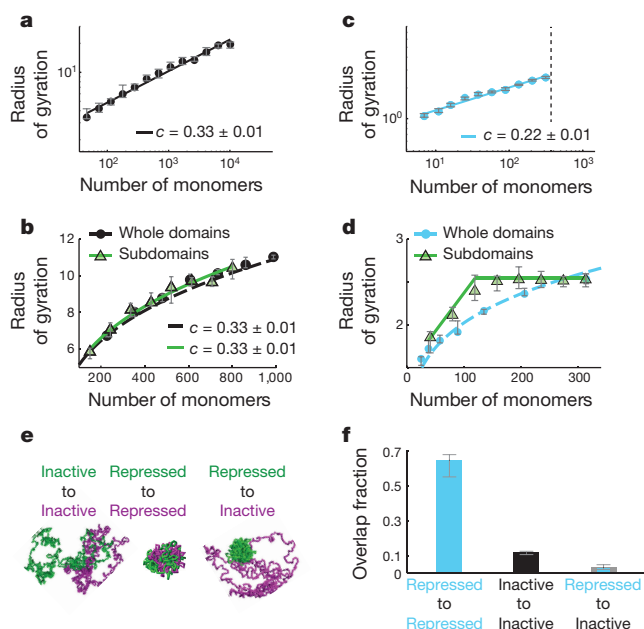


Figure 4 | Computational modelling of chromatin packaging for Inactive and Repressed domains. **a**, Radius of gyration of polymer domains as a function of the domain length for simulated polymers confined in a small volume to emulate Inactive chromatin. The line indicates the power-law fit with the scaling exponent $c = 0.33$. **b**, Radius of gyration of subdomains of a parent Inactive domain (green triangles) compared to the whole-domain scaling data (black circles) with power-law fits and scaling exponents shown. **c**, Radius of gyration as a function of the domain length for the simulated sticky polymer domain, embedded in a non-stick, confined polymer, to emulate Repressed chromatin. The light blue line represent the power-law fit with the scaling exponent $c = 0.22$. The black dotted line indicates the point at which the closest possible packing of monomers is reached, causing c to deviate from 0.22 and approach 0.33 beyond this point. The exponent c also deviates from 0.22 at the small monomer number end because such a short polymer cannot sufficiently bend. **d**, Radius of gyration of the subdomains (green triangles) of the parent sticky domain in comparison with the whole-domain scaling data (light blue circles). **e**, Snapshots of simulations showing adjacent subdomains in the Inactive (non-sticky) chromatin (left), adjacent subdomains of the Repressed (sticky) chromatin (middle), and adjacent Repressed and Inactive chromatin domains (right). **f**, Quantification of the overlap fraction (Supplementary Methods) between the adjacent polymer regions illustrated in **e**. Error bars indicate 95% confidence intervals derived from resampling ($n \approx 20$ simulations).

To test if these structural features for Repressed domains indeed depend on PcG proteins, we used RNA interference (RNAi) to knockdown a Polycomb repressive complex I (PRC1) component, Polyhomeotic (Ph), which contains a sterile alpha motif that is capable of self-interaction²⁹. Quantitative PCR (qPCR) experiments indicated an $\sim 80\%$ knockdown efficiency of the two copies of *ph*, *ph-p* and *ph-d* (Extended Data Fig. 7a). Functionally, Ph-knockdown led to a substantial increase in expression of many known PcG-target genes, but did not affect the expression of the hundreds of genes from the Inactive and Active domains that we imaged (Extended Data Fig. 7a, b), consistent with previous work showing the important role of PcG proteins in transcriptional repression^{2,21–23}. Structurally, Ph-knockdown abolished all of the three unique packaging properties that we observed for the Repressed domains. First, Ph-knockdown caused substantial decompaction of the Repressed domains to an extent that their 3D sizes approached those of Active domains, and the power-law scaling exponents of the Repressed domains increased substantially such that the domains no longer became more densely packed with increasing domain length (Fig. 1c, d, hollow circles, Extended Data Fig. 7c). Second, Ph-knockdown abolished the saturation effect that we

observed for the subdomains of the Repressed domains (Extended Data Fig. 7d). Finally, the spatial overlap between neighbouring Repressed and Active domains in Ph-knockdown cells increased markedly to 20–30% (Fig. 3b, c), becoming similar to the extent of overlap observed between neighbouring Active subdomains. The increase in intermixing of Repressed and Active domains upon Ph-knockdown is statistically highly significant (Wilcoxon test $P = 1 \times 10^{-18}$, $n \approx 150$ cells). These data indicate that the unique structural properties that we observed for the Repressed domains indeed depend on PcG proteins, potentially through the self-interactions of the Ph component of PRC1.

Our experimental data also revealed significant locus-specific variations in chromatin packaging, particularly for the Active domains (Extended Data Fig. 8). After normalization for different domain lengths based on our observed scaling law, 42% of the pairs of examined Active domains showed statistically distinct degrees of compaction ($P < 0.05$), while this fraction was only 16% for the examined Repressed domains (Extended Data Fig. 8b). These variations may be linked to local genomic context, such as different activating histone modifications or transcription factor binding²⁰, or different levels of transcriptional activity. Indeed, our analysis suggested that domains with higher transcriptional activity (measured by RNA-seq) appeared to have less dense packaging, as did domains with higher than average density of CTCF binding, although the degrees of correlation were only moderate (Extended Data Fig. 9).

Taken together, our results showed that direct visualization of DNA in the nucleus with nanometre-scale resolution can reveal previously unseen structural details of chromatin. We observed three distinct spatial organizations for chromatin in *Drosophila* cells, each corresponding to a different epigenetic state. Transcriptionally inactive chromatin, lacking substantial post-translational modification or transcription factor binding, adopts a 3D organization that resembles the fractal globule state of a polymer. On the other hand, Polycomb-repressed chromatin adopts a substantially more compact packaging configuration. Within these Repressed domains, chromatin is highly intermixed, much more so than within the constitutively inactive chromatin. These Polycomb-repressed domains harbour genes encoding key developmental transcription factors, whose misexpression can have detrimental consequences in differentiated cells. Such a high-density packaging and high degree of chromatin intermixing may contribute to the inaccessibility of DNA to the transcription machinery, and thereby may help ensure that transcription is repressed in this chromatin state^{21,22}. Moreover, our observation that Polycomb-repressed domains become increasingly more compact as the domain length increases could explain why PcG-targeted genes (such as the well-known *Hox* genes) tend to cluster in the genome. Concatenating PcG domains would increase domain length and therefore could lead to more compact packaging and potentially a more complete repression of gene expression. Notably, these Polycomb-repressed domains also strongly exclude intermixing with nearby (Active) chromatin, which may help prevent “bystander” activation³⁰ caused by transcription activators and enhancers accidentally brought into proximity of PcG-targeted genes by the Active chromatin. We note that the Polycomb-repressed domains investigated here all have relatively high density of bound PcG proteins (Supplementary Methods); it is thus possible that the structural properties that we observed for Repressed domains may not apply to all Polycomb-enriched domains and that a sufficiently high PcG density is needed to establish these properties. Finally, we observed a far less compact packaging for transcriptionally active domains as compared to transcriptionally inactive domains; such an open structure may help transcription factors and polymerase machinery gain access to the encoded genes and facilitate transcription. Thus, our data suggest that epigenetic factors not only partition the genome into distinct, one-dimensional domains of different biochemical properties, but can also shape the three-dimensional nanoscale structures of these domains, which may work in concert with the biochemical properties to regulate gene expression.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 June; accepted 26 November 2015.

Published online 13 January 2016.

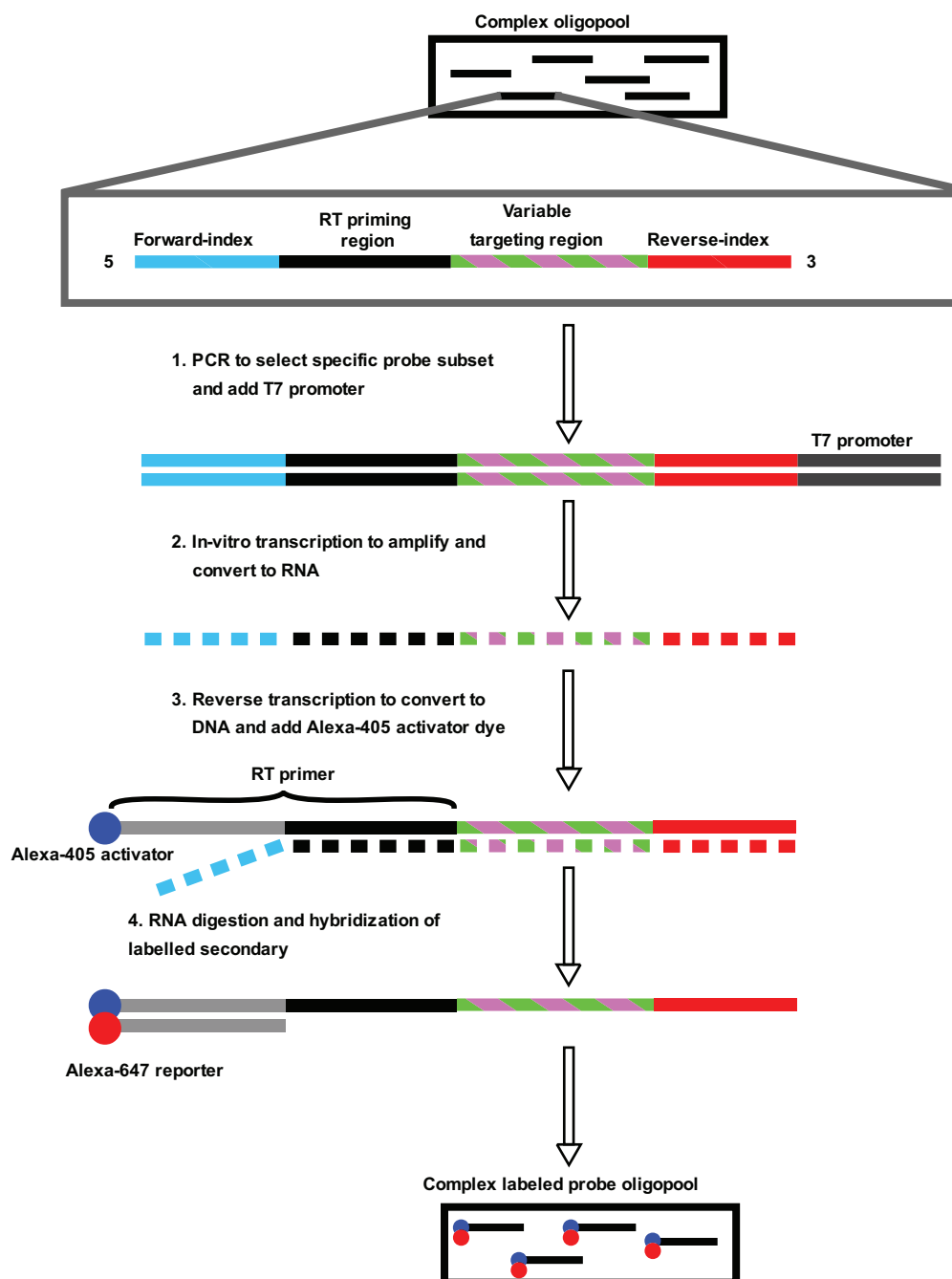
- Kornberg, R. D. & Lorch, Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**, 285–294 (1999).
- Bickmore, W. A. The spatial organization of the human genome. *Annu. Rev. Genomics Hum. Genet.* **14**, 67–84 (2013).
- Gibcus, J. H. & Dekker, J. The hierarchy of the 3D genome. *Mol. Cell* **49**, 773–782 (2013).
- Levine, M., Cattoglio, C. & Tjian, R. Looping back to leap forward: transcription enters a new era. *Cell* **157**, 13–25 (2014).
- Cremer, T. & Cremer, M. Chromosome territories. *Cold Spring Harb. Perspect. Biol.* **2**, a003889 (2010).
- Gorkin, D. U., Leung, D. & Ren, B. The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell* **14**, 762–775 (2014).
- Sexton, T. & Cavalli, G. The role of chromosome domains in shaping the functional genome. *Cell* **160**, 1049–1059 (2015).
- Galupa, R. & Heard, E. X-chromosome inactivation: new insights into *cis* and *trans* regulation. *Curr. Opin. Genet. Dev.* **31**, 57–66 (2015).
- Bernstein, B. E., Meissner, A. & Lander, E. S. The mammalian epigenome. *Cell* **128**, 669–681 (2007).
- Bickmore, W. A. & van Steensel, B. Genome architecture: domain organization of interphase chromosomes. *Cell* **152**, 1270–1284 (2013).
- Rivera, C. M. & Ren, B. Mapping human epigenomes. *Cell* **155**, 39–55 (2013).
- Gómez-Díaz, E. & Corces, V. G. Architectural proteins: regulators of 3D genome organization in cell fate. *Trends Cell Biol.* **24**, 703–711 (2014).
- Beliveau, B. J. *et al.* Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl Acad. Sci. USA* **109**, 21301–21306 (2012).
- Beliveau, B. J. *et al.* Single-molecule super-resolution imaging of chromosomes and *in situ* haplotype visualization using Oligopaint FISH probes. *Nat. Commun.* **6**, 7147 (2015).
- Boyle, S., Rodesch, M. J., Halvensleben, H. A., Jeddah, J. A. & Bickmore, W. A. Fluorescence *in situ* hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis. *Chromosome Res.* **19**, 901–909 (2011).
- Yamada, N. A. *et al.* Visualization of fine-scale genomic structure by oligonucleotide-based high-resolution FISH. *Cytogenet. Genome Res.* **132**, 248–254 (2011).
- Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, 412 (2015).
- Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods* **3**, 793–796 (2006).
- Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, 810–813 (2008).
- Filion, G. J. *et al.* Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143**, 212–224 (2010).
- Simon, J. A. & Kingston, R. E. Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol. Cell* **49**, 808–824 (2013).
- Grossniklaus, U. & Paro, R. Transcriptional silencing by polycomb-group proteins. *Cold Spring Harb. Perspect. Biol.* **6**, a019331 (2014).
- Cheutin, T. & Cavalli, G. Polycomb silencing: from linear chromatin domains to 3D chromosome folding. *Curr. Opin. Genet. Dev.* **25**, 30–37 (2014).
- Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
- Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
- Mirny, L. A. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res.* **19**, 37–51 (2011).
- Halverson, J. D., Smrek, J., Kremer, K. & Grosberg, A. Y. From a melt of rings to chromosome territories: the role of topological constraints in genome folding. *Rep. Prog. Phys.* **77**, 022601 (2014).
- Nicodemi, M. & Pombo, A. Models of chromosome structure. *Curr. Opin. Cell Biol.* **28**, 90–95 (2014).
- Isono, K. *et al.* SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev. Cell* **26**, 565–577 (2013).
- Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nature Rev. Genet.* **13**, 613–626 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank H. Babcock for help with instrumentation and technical advice, J. Kondev for discussions about the data, A. Goloborodko for sharing custom OpenMM force-field, and S. Nguyen for discussions on RNAi knockdown protocol. This work was supported in part by the National Institutes of Health (X.Z., C.W., L.A.M.). A.N.B. acknowledges support by the Damon Runyon Foundation postdoctoral fellowship. J.R.M. acknowledges support of the Helen Hay Whitney Foundation postdoctoral fellowship. S.W. acknowledges support of the Jane Coffin Childs Foundation postdoctoral fellowship. X.Z. is a Howard Hughes Medical Institute Investigator.

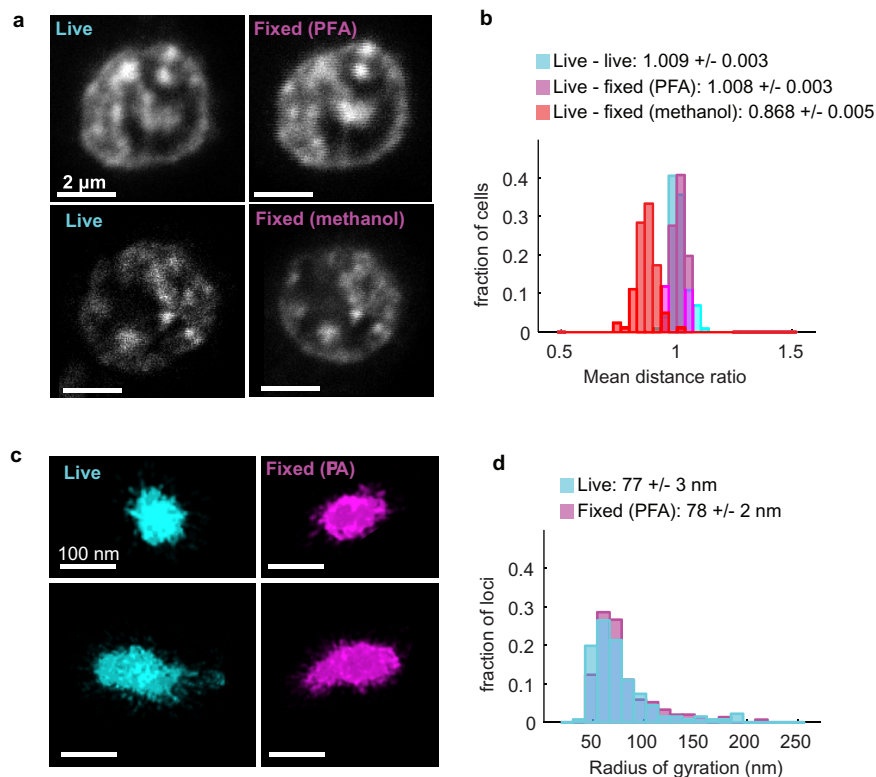
Author Contributions A.N.B. and X.Z. conceived the project. A.N.B., B.B., and X.Z. designed the experiments and simulations with input from C.-t.W. and L.A.M.; A.N.B. and B.B. performed the experiments, data analysis and simulations. J.R.M. conceived the high yield probe synthesis method, and J.R.M. and A.N.B. developed this method. B.J.B. assisted with initial probe synthesis. S.W. assisted with live-cell STORM experiments. G.F. and M.I. assisted with methods for simulation. A.N.B., B.B., L.A.M., C.-t.W. and X.Z. interpreted the results. A.N.B., B.B., J.R.M. and X.Z. wrote the manuscript, with inputs from B.J.B., G.F., M.I., L.A.M. and C.-t.W.

Author Information RNA-Seq data have been deposited at Gene Expression Omnibus (GEO) under the accession code GSE75060. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to X.Z. (zhuang@chemistry.harvard.edu).



Extended Data Figure 1 | Schematic of oligonucleotide probe design and synthesis. A unique pair of index primers are used in a PCR reaction to selectively amplify the templates for the probe set of interest from a complex pool of custom, array-derived oligonucleotides. These templates are then amplified and converted to RNA in an *in vitro* transcription reaction. The RNA products are converted back to DNA in a reverse-transcription

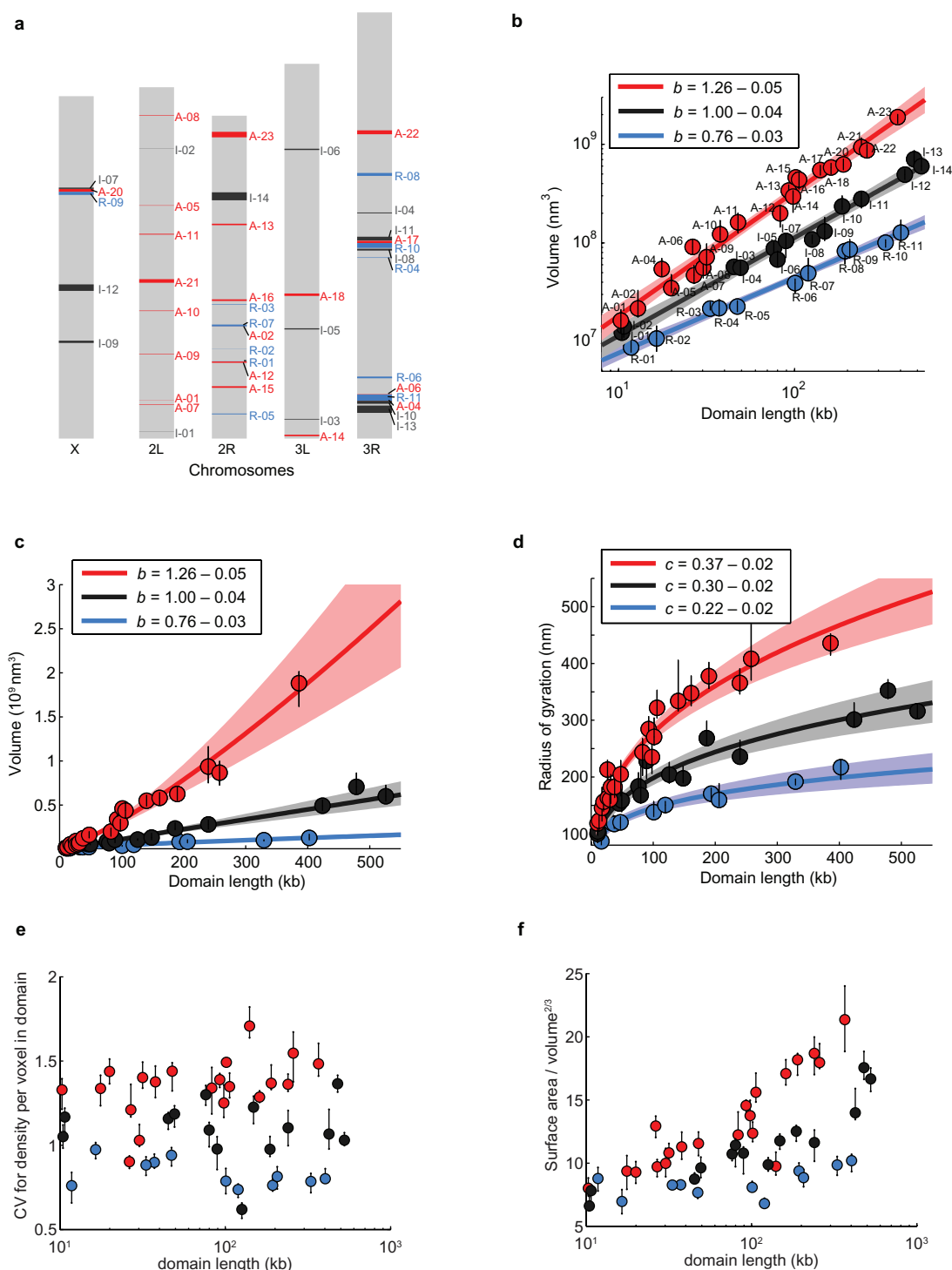
reaction using a primer labelled with an activator dye, Alexa-405, which incorporates the dye into the resulting single-stranded DNA probe. Finally, a 32-nucleotide oligonucleotide attached to Alexa-647 is hybridized to all of the probes. The photoswitchable dye, Alexa-647, is used for STORM imaging. The activator dye, Alexa-405, facilitates the 405-nm light induced reactivation of the Alexa-647 dye.



Extended Data Figure 2 | Effect of cell fixation on chromatin sizes.

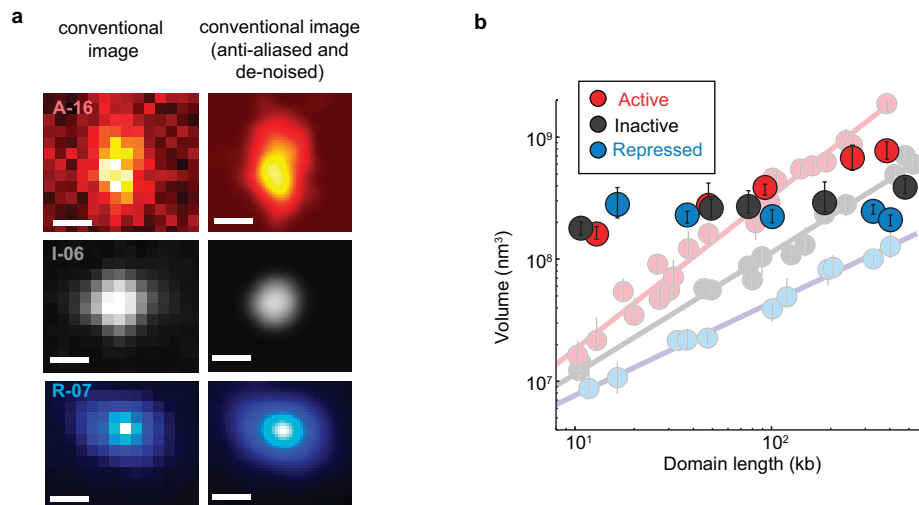
a, Top, example images of DNA in a Kc_{167} cell, visualized with the viable DNA dye Hoechst 33342, both in the live cell before fixation and in the same cell after applying our fixation buffer (osmotically balanced methanol-free formaldehyde in PBS). Bottom, same as the top panels but for a Kc_{167} cell before and after fixation with methanol, a fixative that is known to cause a shrinkage effect. **b**, Quantifications of the distances between chromatin features in live and fixed cells. Corresponding chromatin features were identified in the live and corresponding fixed cell images through scale-invariant feature transform (SIFT) registration. We measured distances between pairs of identified SIFT features in each cell, and calculated the ratio between the median inter-feature distances before and after fixation for each cell. Plotted here are the histograms of ratios determined from many cells for fixation with our osmotically balanced methanol-free formaldehyde fixation buffer (magenta), for a 'mock fixed' condition in which the growth media was replaced with fresh media without any fixation reagent (cyan), and for fixation with methanol

(red), $n \approx 80$ cells in each cases. The average ratios are 1.009 ± 0.003 and 1.008 ± 0.003 for fixation with our fixation buffer and the mock fixation, respectively, indicating a lack of shrinkage effect. In contrast, the average ratio for the methanol-fixation case (0.868 ± 0.005) is appreciably less than one, indicating a chromosome shrinkage induced by methanol. **c**, STORM images of TRF1-mMaple3 labelled telomeres in live and fixed HEK293 cells. Cells are fixed with our osmotically balanced methanol-free formaldehyde in PBS. Two examples of telomere STORM images are shown for each condition. **d**, Quantifications of the radius of gyration of the telomeric domains in live and fixed cells. We determined the radius of gyration for each telomere structure and plotted here are the histograms of the radii of gyration across about 150 telomeres from around 30 cells for live (cyan) and fixed (magenta) cells. The average radius of gyration is 77 ± 3 nm for live cells and 78 ± 2 nm for fixed cells, again indicating that there is no significant chromatin shrinkage effect upon fixation. The telomere size measurement is not limited by our image resolution with mMaple3 (~ 30 nm).



Extended Data Figure 3 | Volume, radius of gyration and other shape characteristics for chromatin domains of various domain lengths in three different epigenetic states. **a**, Scheme of *Drosophila* chromosomes (X, 2L, 2R, 3L, and 3R) with the position of the imaged epigenetic domains marked (red, Active domains A-01 to A-23; black, Inactive domains I-01 to I-14; blue, Repressed domains R-01 to R-11). **b**, log-log plot of the median domain volume as a function of domain contour length reproduced from Fig. 1c but with the domain ID labelled. **c**, As in Fig. 1c but plotted on a linear-linear scale. **d**, Linear plot of the median radius of gyration as a function of domain contour length. **e**, Coefficient of variation (CV) in

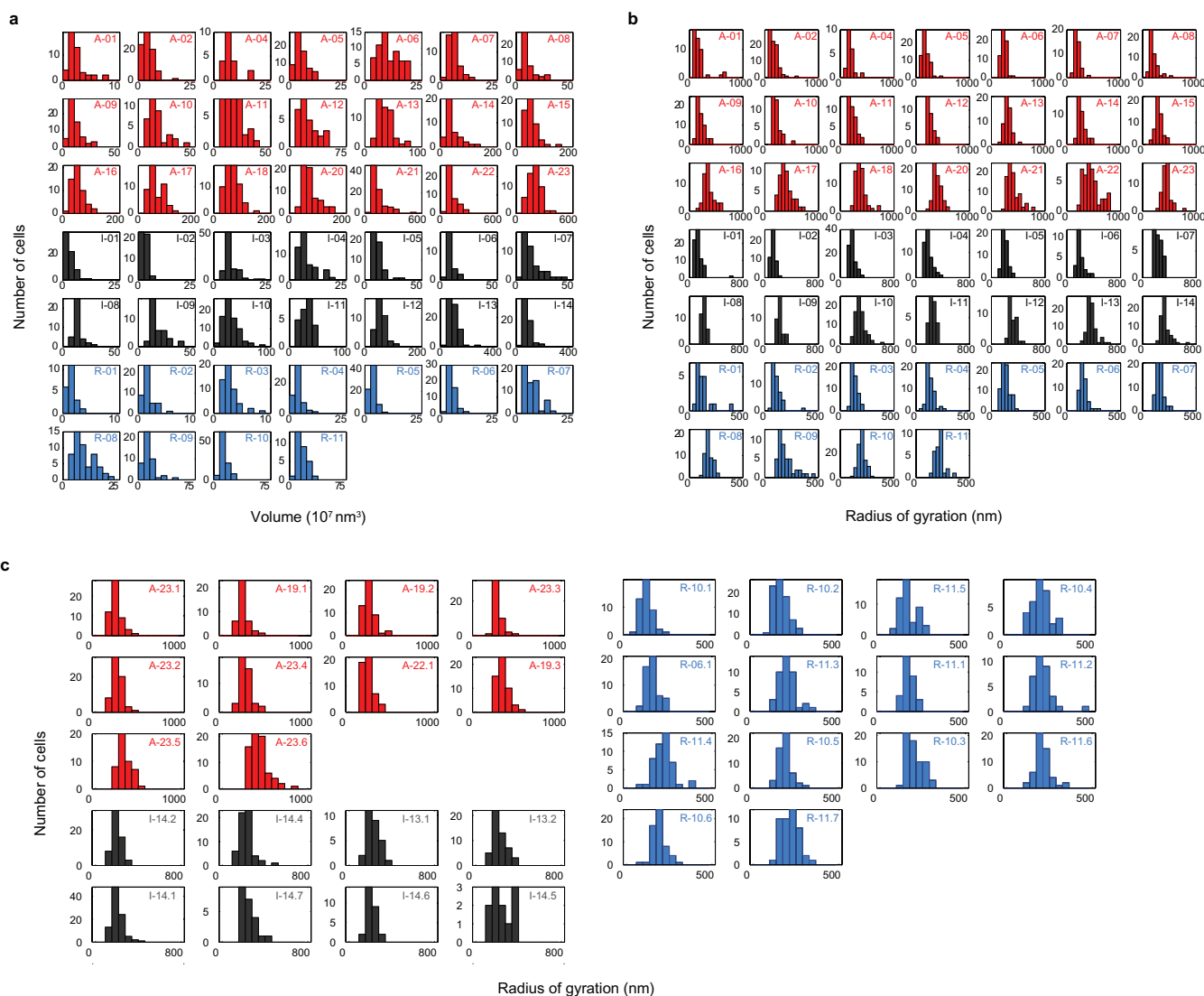
density per voxel for all domains as a function of domain length. CV in density is defined as the ratio of the standard deviation of density to the average density within the domain-occupied volume, which characterizes how uniformly the chromatin is distributed in space within these domains (Supplementary Methods). **f**, Ratio of surface area to volume $^{2/3}$ for all domains as a function of domain length. This surface-to-volume parameter characterizes the complexity of the physical shapes taken by the domains in 3D (Supplementary Methods). Error bars represent 95% confidence intervals derived from resampling ($n \approx 50$ cells).



Extended Data Figure 4 | Conventional images of chromatin domains and domain volume characterization based on conventional images.

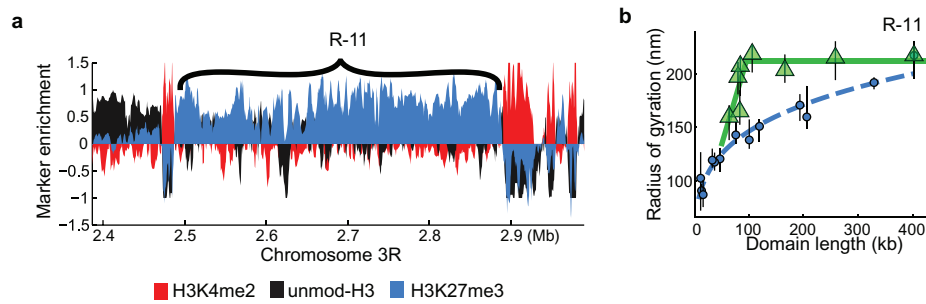
a, Blow-up view of the conventional images of chromatin domains shown in Fig. 1b. The left column shows the raw conventional, wide-field images, with pixel size defined by our camera. The right column shows the corresponding anti-aliased and de-noised images. **b**, Quantification of the median domain volume determined from conventional images (foreground symbols), overlaid on the median volume determined from

STORM data plotted in Fig. 1c (faint background symbols and lines). Error bars represent 95% confidence intervals derived from resampling ($n \approx 50$ cells). Note that the conventional images may not only cause an artificial increase in domain size, especially severe for those domains whose physical sizes are smaller than the image resolution, but can also lead to an apparent decrease in domain size in some cases when the thin protrusions were too dim to detect by conventional imaging.



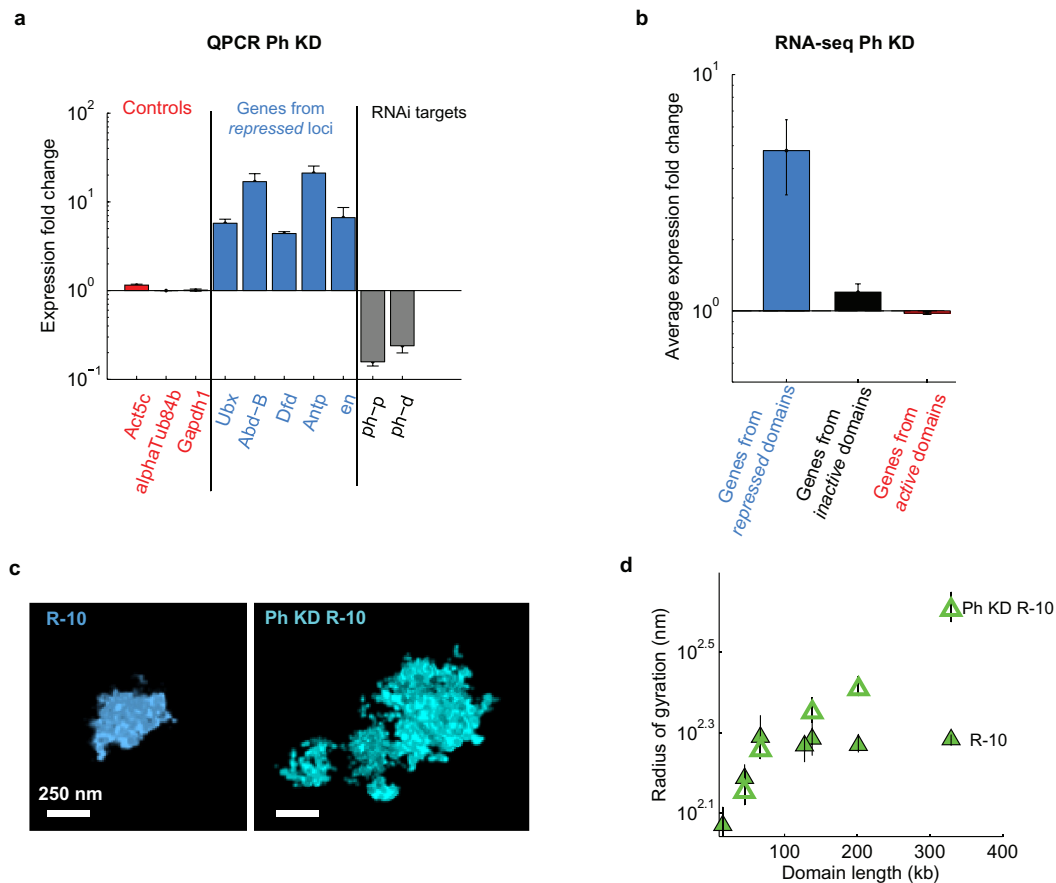
Extended Data Figure 5 | Distributions of domain volume and radius of gyration of different epigenetic domains and subdomains over all imaged cells. a, Histograms of domain volume for all imaged cells for each of the domains shown in Extended Data Fig. 2a. Red, Active domains; black, Inactive domains; light blue, Repressed domains. The domain IDs are indicated in the upper right corner of each plot. The x -axis (volume)

range has been adjusted for each domain to ensure the readability of the histogram. **b,** Histograms of the radius of gyration for each of the imaged domains in all cells. **c,** Histograms of the radius of gyration for subdomains of Active (red), Inactive (black) and Repressed (blue) chromatin, shown in Fig. 2b and Extended Data Fig. 6, for all imaged cells. The subdomain IDs are indicated in the upper right corner of each plot.



Extended Data Figure 6 | Additional data on the scaling behaviour of subdomains of Repressed chromatin. **a**, Enrichment profile of H3K4me2 (red), H3K27me3 (light blue) and unmodified H3 (black) in a genomic region harbouring the Repressed domain R-11 (Antennapedia complex). **b**, The radius of gyration of subdomains (green triangles) of R-11 as a

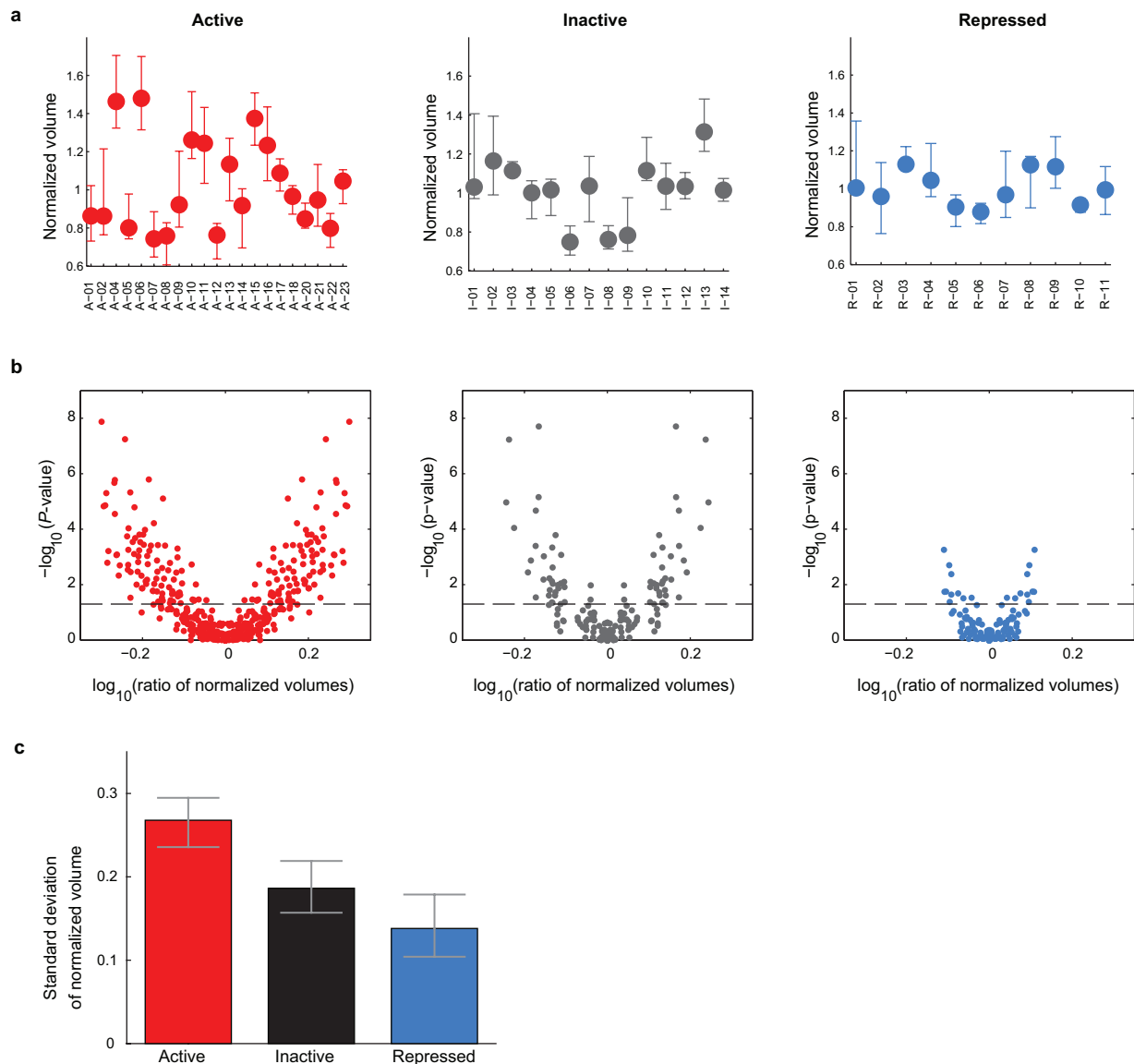
function of subdomain length compared to the scaling of whole Repressed domains (light blue circles). The data shown in Fig. 2b, right panel are for the R-10 domain (Bithorax complex). The light blue dashed line indicates the power-law fit for the whole domain data. Green lines are to guide the eye. Error bars represent 95% confidence intervals ($n \approx 50$ cells).



Extended Data Figure 7 | Effect of polyhomeotic (Ph) knockdown.

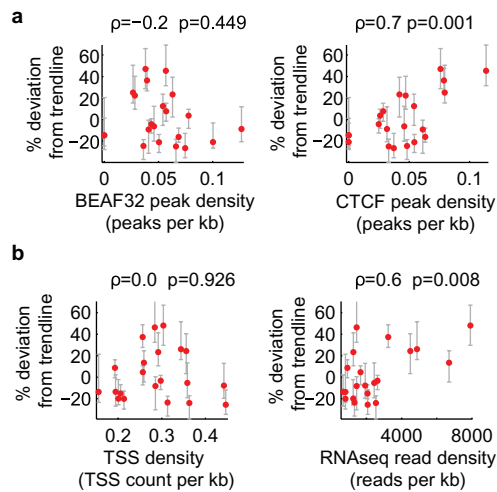
a, Quantification of relative change in gene expression by qPCR (mean \pm s.e.m., $n = 3$ biological replicates) upon *ph-p* and *ph-d* double knockdown. Grey bars, expression fold change of *ph-p* and *ph-d* upon double knockdown. Light blue bars, expression fold change of five Polycomb target genes, *Ubx*, *Abd-B*, *Dfd*, *Antp*, *en*. Red bars, expression fold change of three control genes, *Act5c*, *alphaTub84b* and *Gapdh1*, that are not targeted by Polycomb. Expression fold change was determined as the ratio between the signal detected in *ph-p* and *ph-d* double knockdown cells and that detected in wild-type cells. **b**, Average expression fold change upon *ph-p* and *ph-d* double knockdown for all genes in all of the Active (red), Inactive (black) and Repressed (light blue) domains included in our study. The expression fold change is defined as the ratio of expression level

measured in Ph-knockdown cells to that measured in the wild-type control cells determined by next generation RNA sequencing (mean \pm s.e.m., $n = 45$, 89 and 532 genes for Repressed, Inactive and Active domains, respectively, 2 biological replicates). Expression level was measured in units of read fragments per kilobase per million reads (FPKM). Note, some genes (9 from Repressed regions, 11 from Inactive regions and 2 from Active regions) are excluded from the average expression fold change calculation because they received zero counts in the wild-type control cells. **c**, Example images of the R-10 domain (Bithorax complex) in wild-type (left) and Ph-knockdown (right) cells. **d**, Radius of gyration versus domain length for subdomains of R-10 in wild-type cells (solid green triangles) and Ph-knockdown cells (hollow green triangles). Error bars represent 95% confidence intervals ($n \approx 50$ cells).



Extended Data Figure 8 | Locus-to-locus variation observed for the three types of epigenetic domains after normalization based on the observed scaling law over domain length. a, Top, the normalized volume for domains of Active (left), Inactive (middle), and Repressed (right) chromatin. Normalized volume is defined as the ratio of median volume of the domain to the expected volume calculated from the power-law scaling fits shown in Fig. 1c. Error bars represent 78% confidence intervals, such that there is a less than 5% chance that domains with non-overlapping error bars are not distinct ($n \approx 50$ cells). **b**, Volcano plots of the relative differences in volume

between all pairs of Active domains (left), Inactive domains (middle), or Repressed domains (right), after the normalization shown in **a**. Each data point represents one pair of domains with their ratio of the normalized volumes plotted on the x -axis and the P value of their normalized volume difference plotted on the y -axis. The dashed line is at a P value of 0.05. All dots above this line represent pairs of domains in which the normalized volume of one domain is statistically distinguishable from that of the other domain. **c**, Standard deviation of the normalized volumes for each domain type. Error-bars represent 95% confidence intervals ($n \approx 50$ cells).



Extended Data Figure 9 | Additional factors correlating with the domain volume after normalizing the effect of domain length for Active domains. To normalize for the effect of domain length, we determined the percent deviation of the volumes of Active domains from the power-law scaling trend line shown in Fig. 1c, and hereafter refer to this value as percent deviation from trend line. **a**, Correlation of the percent deviation from trend line with the binding density of the insulator proteins BEAF32 (left) and CTCF (right). Binding density was determined from the density of peaks per kb in Dam-ID data²⁰. Peaks were defined as local maxima at least 2 standards deviation above the mean. **b**, As in **a** but for correlation with transcription start site (TSS) density (left) and RNA-seq total read density (right). The TSS density is defined as the average number of TSSs per kb in the domain and the RNA-seq total read density is defined as the total number of reads mapping to the domain measured using RNA sequencing divided by the domain length in kb. Error bars represent 95% confidence intervals ($n \approx 50$ cells). **c**, Pearson correlation coefficients and corresponding P values for the correlation of percent deviation from trend

Factor	Pearson Correlation Coefficient	p-value
TSS density	0.0	0.926
RNA-seq read density	0.6	0.008
Ave. gene expression	0.5	0.011
Max. gene expression	0.1	0.575
CTCF peak density	0.7	0.001
Su(Hw) peak density	-0.4	0.069
BEAF32 peak density	-0.2	0.449

line with the indicated genomic factors. Average gene expression refers to the average expression value in FPKM (read fragments per kilobase per million reads) of all genes in the domain. Maximum gene expression refers to the FPKM of the most highly expressed gene in the domain. Su(Hw) is an insulator protein like BEAF32 and CTCF. We noticed a weak trend in which domains with higher binding densities of the insulators BEAF32 or Su(Hw) are slightly more compact. Although this trend is consistent with the hypothesis that insulator proteins may function as loop forming factors and that loops may lead to more compact domains¹², the correlation detected here was not statistically significant. Further analysis with improved sensitivity in detection of BEAF32 or Su(Hw) binding sites might uncover a stronger effect, so our data do not rule out the insulator loop hypothesis. Similarly, we caution that the positive correlation observed with the density of CTCF binding sites might reflect the preference of CTCF to bind open chromatin regions (such as enhancers and promoters), and does not necessarily suggest that CTCF binding induces a more open chromatin state.

Extended Data Table 1 | Epigenetic domains and subdomains probed

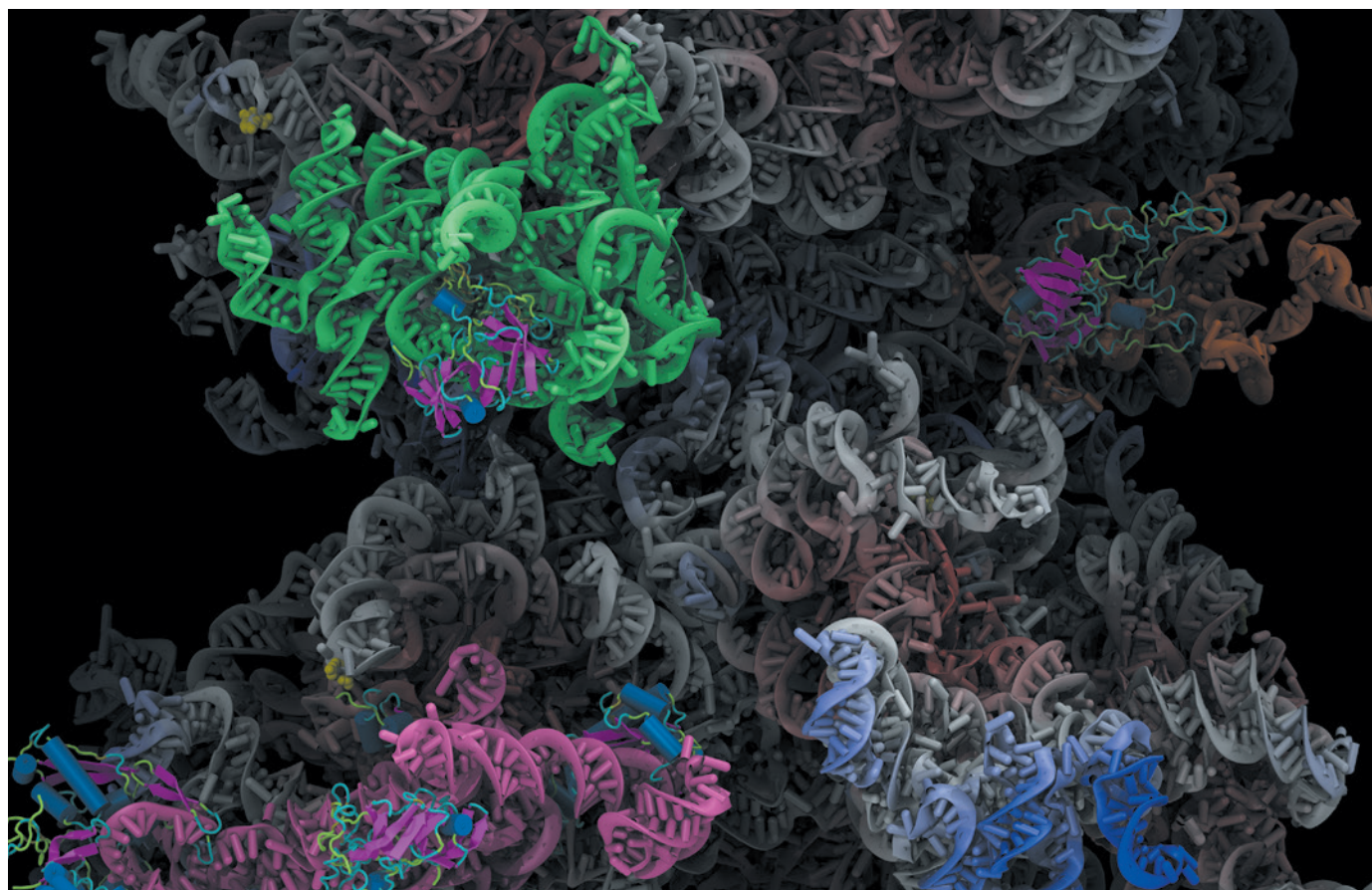
Domains				Subdomains			
ID	Genome coordinate	Length (bp)	Selected genes	ID	Genome coordinate	Length (bp)	Selected genes
A-01	chr2L:2489241-2499536	10295	oaf	A-19.1	chrX:1907754-1977754	70000	(Many genes)
A-02	chr2R:7338881-7351765	12884	E(Pc)	A-19.2	chrX:1977754-2047754	70000	(Many genes)
A-03	chr3R:12464802-12479673	14871	Cctgamma	A-19.3	chrX:1907754-2047754	140000	(Many genes)
A-04	chr3R:2470393-2487998	17605	TafI	A-22.1	chr3R:19929199-20029199	100000	(Many genes)
A-05	chr2L:15254575-15274520	19945	yuri	A-23.1	chr2R:19726615-19787585	60970	(Many genes)
A-06	chr3R:2890667-2916994	26327	alphaTub84b, Dpck, Aly	A-23.2	chr2R:19809874-19888410	78536	(Many genes)
A-07	chr2L:2210964-2237817	26853	mio, Uch, rti,	A-23.3	chr2R:19906491-19976552	70061	(Many genes)
A-08	chr2L:21153308-21183571	30263	Acon, bur, Mpp6, Nbr, Mcm10, Ret	A-23.4	chr2R:20000481-20092780	92299	(Many genes)
A-09	chr2L:5517877-5549514	31637	Lam, Cap-D3, tomb, Oscillin, Hel25E, cl	A-23.5	chr2R:19726615-19888410	161795	(Many genes)
A-10	chr2L:8353587-8391384	37797	Aats-ala, RpS13, CSNS, Pp2A-29B, Rcd4, Mur29b	A-23.6	chr2R:19809874-20092780	282906	(Many genes)
A-11	chr2L:13363559-13411277	47718	Vm34Ca,RpL24, loqs. betaCop, Tap42,Nnp-1	I-13.1	chr3R:1786654-1886654	100000	Pbprp3,Os-E
A-12	chr2R:4972056-5054981	82925	(Many genes)	I-13.2	chr3R:1686653-1786653	100000	(Contains only genes of little known function)
A-13	chr2R:13979742-14072371	92629	(Many genes)	I-14.1	chr2R:15602615-15702615	100000	Or56a,Obp56g
A-14	chr3L:167832-265882	98050	(Many genes)	I-14.2	chr2R:15702616-15802615	99999	Obp56h,Toll-7,Obp56i
A-15	chr2R:3328439-3429969	101530	(Many genes)	I-14.3	chr2R:15802616-15852615	49999	(Contains only genes of little known function)
A-16	chr2R:9019018-9125165	106147	(Many genes)	I-14.4	chr2R:15852616-15952615	99999	(Contains only genes of little known function)
A-17	chr3R:12811417-12951671	140254	(Many genes)	I-14.5	chr2R:15952616-16128463	175847	18w
A-18	chr3L:9344016-9505324	161308	(Many genes)	I-14.6	chr2R:15702616-15852615	149999	Obp56h,Toll-7,Obp56i
A-19	chrX:1907754-2083414	175660	(Many genes)	I-14.7	chr2R:15802616-15952615	149999	(Contains only genes of little known function)
A-20	chrX:16161627-16351359	189732	(Many genes)	I-14.8	chr2R:15852616-16128463	275847	18w
A-21	chr2L:10203488-10443191	239703	(Many genes)	R-06.1	chr3R:3967588-4043731	76143	grn
A-22	chr3R:19929199-20187132	257933	(Many genes)	R-10.1	chr3R:12571412-12586412	15000	(Ubx regulatory region)
A-23	chr2R:19726615-20092780	366165	(Many genes)	R-10.2	chr3R:12571412-12616414	45002	(Ubx regulatory region)
I-01	chr2L:455203-465638	10435	(Contains only genes of little known function)	R-10.3	chr3R:12481406-12619389	137983	Ubx
I-02	chr2L:18991061-19001768	10707	(Contains only genes of little known function)	R-10.4	chr3R:12616227-12683098	66871	abd-A
I-03	chr3L:1250391-1295624	45233	(Contains only genes of little known function)	R-10.5	chr3R:12683230-12810708	127478	Abd-B
I-04	chr3R:14754844-14804163	49319	(Contains only genes of little known function)	R-10.6	chr3R:12481406-12683098	201692	Ubx, abd-A
I-05	chr3L:7154799-7231065	76266	(Contains only genes of little known function)	R-11.1	chr3R:2487143-2570646	83503	zen2, pb, lab
I-06	chr3L:18903046-18983014	79968	(Contains only genes of little known function)	R-11.2	chr3R:2568667-2653159	84492	Dfd, bcd, zen, zen2
I-07	chrX:16351360-16440669	89309	para (DmNav1)	R-11.3	chr3R:2641835-2721984	80149	Scr, ftz
I-08	chr3R:12312771-12438748	125977	Scp2, Decay, Gr89a, npf	R-11.4	chr3R:2721063-2827254	106191	Antp
I-09	chrX:6269580-6417829	148249	(Contains only genes of little known function)	R-11.5	chr3R:2825712-2889707	63995	Sodh-1, (Antp regulatory region)
I-10	chr3R:2287031-2473421	186390	sunz, glob2, dpR-11	R-11.6	chr3R:2487143-2653159	166016	Scr, Dfd, bcd, zen, zen2, pb, lab
I-11	chr3R:12977555-13217431	239876	beat-IIa, beat-IIb, Brf	R-11.7	chr3R:2568667-2827254	258587	Antp, ftz, Scr, Dfd, bcd, zen, zen2
I-12	chrX:9679192-10103054	423862	Yp1, Yp2, Cht6, Ser7, Idgf4				
I-13	chr3R:1686653-2165126	478473	Or83c, Os-E, Osi3-Osi19, Pbprp3, Obp83cd, Obp83ef, Obp83g				
I-14	chr2R:15602615-16128463	525848	18w, Toll-7, Obp56h,Obp56i				
R-01	chr2R:5080631-5092410	11779	unpg				
R-02	chr2R:5860081-5876517	16436	eve				
R-03	chr2R:8770090-8803242	33152	vg				
R-04	chr3R:11833445-11870730	37285	pnr, GATAe				
R-05	chr2R:1593366-1640664	47298	ap				
R-06	chr3R:3953538-4054361	100823	grn				
R-07	chr2R:7350135-7469804	119669	en, inv				
R-08	chr3R:17205633-17399201	193568	slou, lbe, lbl, bap				
R-09	chrX:15953918-16160086	206168	disco, disco-r, Tob				
R-10	chr3R:12481406-12810708	329302	Abd-B, abd-A, Ubx (Bithorax-complex)				
R-11	chr3R:2487143-2889707	402564	Antp, ftz, Scr, Dfd, bcd, zen, zen2, pb, lab (Antenopodia-complex)				

List of the domain IDs, locus coordinates, lengths (in base pairs) and representative harboured genes of all epigenetic domains and subdomains imaged in this study.

FINDING FUNCTION IN MYSTERY TRANSCRIPTS

*Little is known about the function of most long non-coding RNAs.
But a suite of new tools might change that.*

KARISSA Y. SANBONMATSU



Conceptual model of a long non-coding RNA that provides structural scaffolding (grey) while binding to proteins (magenta) using highly structured domains (green, pink and orange).

BY KELLY RAE CHI

In 2013, a group of researchers decided to dig deeper into a human embryonic stem-cell line called H1 — and uncovered some surprises. H1 is one of the best known stem-cell lines, yet the team managed to unearth more than 2,000 previously uncharacterized stretches of RNA¹. What is more, 146 of those were exclusive to human embryonic stem cells, offering tantalizing leads into pluripotency — the ability to become any cell type in the body.

These transcripts had gone unnoticed

because they contain repetitive stretches of code that sequence analysers had tended to filter out.

It was a big blind spot. Other labs had uncovered early evidence of RNAs that are rich in repetitive codes and important in human stem cells. As the researchers, who were based mostly at California's Stanford University, examined their haul, they realized that they had hit on exactly these kinds of RNA. Among their list of 146 RNA sequences, says team member Vittorio Sebastiano, three of the most abundant — which they named HPAT2, HPAT3 and HPAT5 — seemed to be necessary for establishing the

pluripotent cells that develop into a human fetus: those that comprise the 'inner cell mass' of an embryo that has yet to implant in the uterus².

These RNA stretches are examples of long non-coding RNAs (lncRNAs) — sequences at least 200 bases long that do not encode proteins. lncRNAs are present in many different kinds of tissue and are often found in specific spots inside a cell. But most lack a defined function and, until recently, were thought to be little more than transcriptional noise.

That view began to shift as more data rolled in showing that the genomic regions from ►

► which lncRNAs are transcribed are more highly conserved through evolution than was imagined, implying that they had some function. But, to this day, a neat and sensible classification system for lncRNAs remains out of reach. They are still ‘each their own snowflake’, says John Rinn, who discovered lncRNAs as a graduate student about 15 years ago and now runs a lab specializing in the molecules at the Broad Institute of MIT and Harvard in Cambridge, Massachusetts.

Now, revolutionary tools such as the genome-editing platform CRISPR–Cas9 are making the task of discovering what individual lncRNAs do much easier.

Some lncRNAs are thought to act as scaffolds for proteins to hang off while they manipulate the packaging of DNA. The functions of others are merely hinted at by their co-occurrence with proteins (see ‘Guilt by association’) or by the effects of their absence — cancers start to spread to other parts of the body, and developmental disorders such as autism spectrum disorder arise.

Today, it is broadly assumed that the molecules do have biological functions, says geneticist John Mattick, who heads the Garvan Institute of Medical Research in Sydney, Australia. “The evidence is on the table. That’s been the big change — and it’s palpable,” he says.

There are still doubters, but arguments now tend to caution against assuming function rather than dismiss its likelihood altogether. In the cautious camp is Michael McManus at the University of California, San Francisco. In 2013, his group mined published data sets of RNA sequences and found tens of thousands of new human lncRNAs, although many were present in cells only in very low amounts³. Most of these still need to be followed up, McManus says. And even with the latest tools, figuring out what the myriad lncRNAs do will be a slog requiring an army of experts.

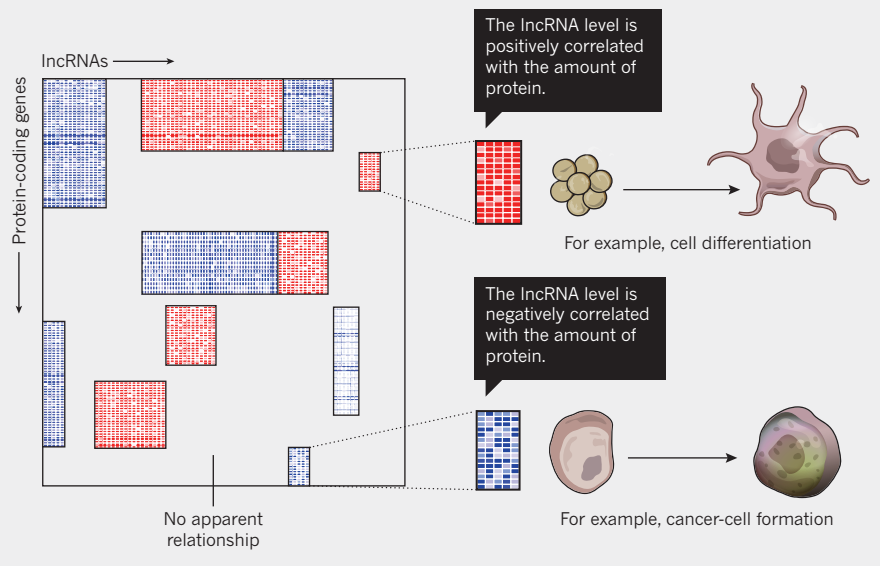
FUNCTION FIRST

Because the list of lncRNAs is so long, a key step is deciding which ones to prioritize for study. Rinn advocates starting with those from regions of the genome that have been already linked to disease. Another idea is to look at where the lncRNAs are located — finding one near a transcription start site might mean that it is involved in regulating the nearby gene, for example. These days, researchers can track the location of molecules inside cells with relative ease. Rinn, along with others at the Broad Institute and at the University of Pennsylvania in Philadelphia, has managed to discern the positions of 61 lncRNA molecules within skin, lung and cervical tumour cells using a technique known as single-molecule RNA-FISH (RNA fluorescence *in situ* hybridization)⁴.

Scientists can also now test lncRNA function by using CRISPR–Cas9 and other gene-editing techniques to interfere with part of the DNA sequence from which it is transcribed or

GUILT BY ASSOCIATION

Grouping long non-coding RNAs (lncRNAs) according to how their presence correlates with the expression of certain proteins can offer clues to what cell processes these mysterious transcripts are involved in.



with the promoter that directs its transcription. Some of these technologies allow labs to quickly screen huge numbers of lncRNAs. The logic is the same as when CRISPR–Cas9 is used to look at the function of a protein-coding gene: introduce single-base deletions or substitutions into the DNA and watch the effects of the altered transcript.

The only problem is that lncRNAs are less likely than proteins to be disabled by subtle alterations, says cancer biologist Howard Chang of Stanford School of Medicine, who develops and applies new methods for studying how lncRNAs work. The sequence alterations often need to be more drastic.

This is where RNA researchers have made CRISPR–Cas9 their own. They have expanded the CRISPR toolbox to include ways to block or prompt the transcription of a specific lncRNA. Rinn and his team have developed yet another approach: a tool known as CRISPR-Display (or ‘CRISP-Disp’). Rinn compares it to a drone that can deliver an item — in this case, a specific lncRNA — anywhere in a cell. If a role in gene activation is suspected because a lncRNA normally lies alongside a certain part of the genome, then that role can be tested by moving the lncRNA to a different genomic location and watching for gene activation in the new spot.

His group had been trying for years to make this happen using older genome-editing methods. Then, once CRISPR–Cas9’s crystal structure was published in 2014 (ref. 5), the team was able to tweak CRISPR’s machinery to carry large packages, and had CRISP-Disp up and running within months. “It’s very high-throughput: we can put 100 different lncRNAs at different sites and ask what they do at once,” says Rinn.

But figuring out function using this and other

CRISPR techniques that block lncRNAs can be more complicated than it seems. For some lncRNAs, the DNA code overlaps with regions that are important for protein-coding genes, so destroying those makes the effects tricky to interpret, says Andrew Bassett, a genome-editing specialist at the University of Oxford, UK. And just because a functional change isn’t seen doesn’t mean that there is no function; the effect may be very subtle, or perhaps revealed only when the cell is faced with a particular threat.

Two often-cited examples of this involve lncRNAs known as NEAT1 and MALAT1. They are abundant in cells and are highly conserved across mammals. Researchers know that they bind DNA to protein, but deleting the DNA stretches has no observable effects in mice. It is a familiar story to researchers in the field. “There are mysteries everywhere,” says Mattick.

THE RNA INTERACTOME

An entirely different approach is to find out what the lncRNAs are interacting with. “It’s still believed, despite the importance of lncRNAs, that, ultimately, they can’t carry out their functions without accessory factors,” says molecular biologist Jeannie Lee of Massachusetts General Hospital in Boston and co-founder of the company RaNA Therapeutics in Cambridge, Massachusetts. These accessory factors are almost always proteins.

Lee and others have set about unveiling lncRNA interactions using a lncRNA — called Xist — that is known to be necessary for silencing one of the two X chromosomes in the cells of female mammals to stop females from having twice as many X-chromosome gene products as males. The proteins that bind to Xist silence gene expression through multiple mechanisms.

But in the past year, scientists have finally made inroads into pinning down the identities of these partners. Now, the proteins are known to not only pull in other molecules that silence transcription, but also to repel some that promote it.

A host of techniques are available for probing a lncRNA's crowd of protein partners: broadly, researchers link RNA and protein together using agents such as formaldehyde or ultraviolet (UV) light, then use mass spectrometry to parse what is bound to what and come up with the 'interactome'. Often, these analyses have many steps, and therefore require the scientist to make many strategic choices. How should the RNA and protein be linked? How can real signals of interactions be distinguished from artefacts? And hanging over all such studies is the problem that RNA often behaves differently *in vitro* from how it does inside a cell.

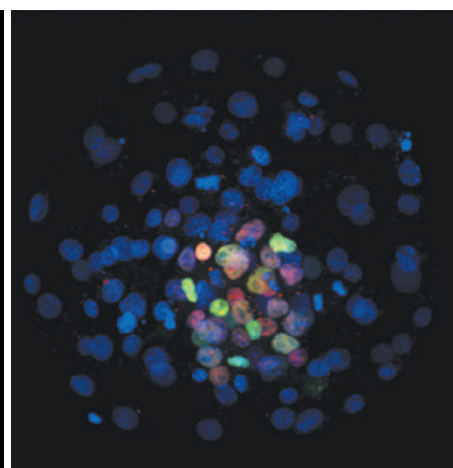
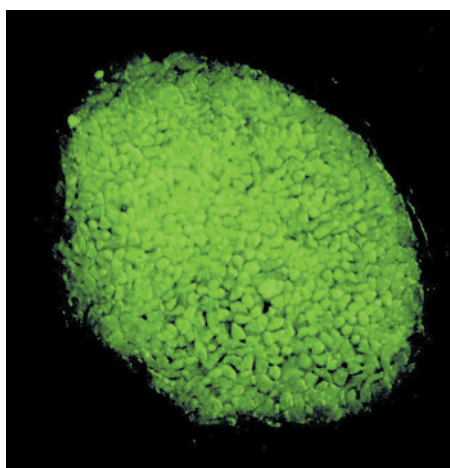
This is why those working on Xist interactions have focused on techniques for identifying RNAs bound to proteins inside living cells. Helping them are recent improvements in the sensitivity of mass spectrometry. In the past year, Chang's group has combined an assay that uses formaldehyde as a linking agent with the latest mass-spectrometry techniques to demonstrate that Xist binds to 81 proteins *in vivo*⁶. Guttman's group used UV light instead — and revealed ten partners, including one not found by Chang's group⁷.

Lee has been working on another UV light method called iDRiP (identification of direct RNA-interacting proteins) and has revealed an Xist interactome of about 100 proteins⁸. That is a complex that rivals the ribosome for size. Lee thinks that iDRiP could be used to look at other lncRNAs but that the protocol would probably need tweaking for each lncRNA. The chosen linking method, she adds, will depend on a number of factors, including where the lncRNA is in the cell and how abundant it is.

Even for the well-studied Xist, the full interactome is far from settled. One of the central debates concerns whether, and how, Xist interacts with a structure called polycomb repressive complex 2 (PRC2), which influences gene expression by modifying proteins called histones. Some groups, including Guttman's, have found little evidence that Xist binds to PRC2; others insist that it does. Guttman thinks that the reason could be down to different experimental protocols: "Xist may bind more strongly *in vitro* than *in vivo*. It's also a question of how one separates the strongest binding interactions from background interactions," he says.

The PRC2 debate highlights the importance of following up interactome assays with tests of whether breaking a lncRNA-protein interaction changes the RNA's function. Guttman

"There are so many different lncRNAs that there's going to be a large zoo of different classes and motifs."



Stem cells known as H1 cells (coloured green, left image) have the ability to develop into any cell type partly because they contain long, non-coding RNAs. These same RNAs are found in the inner cell mass (various colours, right image) of blastomeres.

has found that PRC2 deletion does not seem to affect Xist's ability to silence the X chromosome. By contrast, he says, perturbing the interaction between Xist and another protein implicated in gene silencing — called SHARP — does.

SECRETS IN STRUCTURE

A third avenue for probing the function of lncRNAs is to study their structure. This doesn't predict function as directly as it often does for proteins, but knowing more about an RNA's arches and folds is likely to inform nonetheless. "It's a wide open field that needs a lot of work," says structural biologist Karissa Sanbonmatsu of the Los Alamos National Laboratory in New Mexico. "There are so many different lncRNAs that there's going to be a large zoo of different classes and motifs."

Methods for establishing the secondary structure of a lncRNA include chemical probing strategies, such as one called SHAPE. It involves attaching acetyl groups to the RNA, modifying its backbone only at flexible regions. The modified sites block the enzyme that 'reads' RNA to create a complementary DNA sequence so that short DNA fragments are generated rather than long strands. The fragments can then be sequenced or sized on a gel.

Sanbonmatsu's group was the first to describe, in 2012, the secondary structure of a human lncRNA: the steroid receptor RNA activator (SRA), which had been known for more than a decade to associate with oestrogen receptors⁹.

By chemically probing the bound and unbound SRA, as well as its various domains, Sanbonmatsu revealed the lncRNA in its full glory, including all of its stems, loops and bulges. It looked a lot like the 16s ribosomal RNA, a highly conserved molecular machine. Sanbonmatsu's team has since found other strongly structured lncRNAs, but it is unclear whether most lncRNAs are like this or whether they are floppy, or somewhere in between, she says.

The structural approach, too, has to cope with the problem that RNA behaves differently in test

tubes and cells. And as with binding studies, the latest techniques are being done *in vivo*. In 2012, a team that included Chang described a version of SHAPE that can work inside living cells¹⁰ and has since improved it to characterize thousands of RNA structures simultaneously.

Structural studies, like the others, require a large time investment — so careful choices must be made to focus on the lncRNAs that are most likely to have functions. Luckily, researchers are getting better at such triage, Sanbonmatsu says. She suggests to determine the likelihood of functional significance, scientists should start with lncRNAs that have known phenotypes, then chemically probe them to obtain secondary structures and check the extent to which they are conserved across other species.

Sebastiano already has those boxes ticked for the three lncRNAs that seem to be key in establishing the pluripotent inner cell mass of human embryos: HPAT2, HPAT3 and HPAT5. Now he plans to delve further into the mechanistic details of these factors and has a raft of planned experiments on his list, including assays to ascertain their interactomes as well as structural analyses. "There's a ton of work to do, and this is just the beginning," he says. "But considering that these sequences may explain a lot of our unique features as humans and as primates, the effort is well worth it." ■

Kelly Rae Chi is a freelance science writer based in Cary, North Carolina.

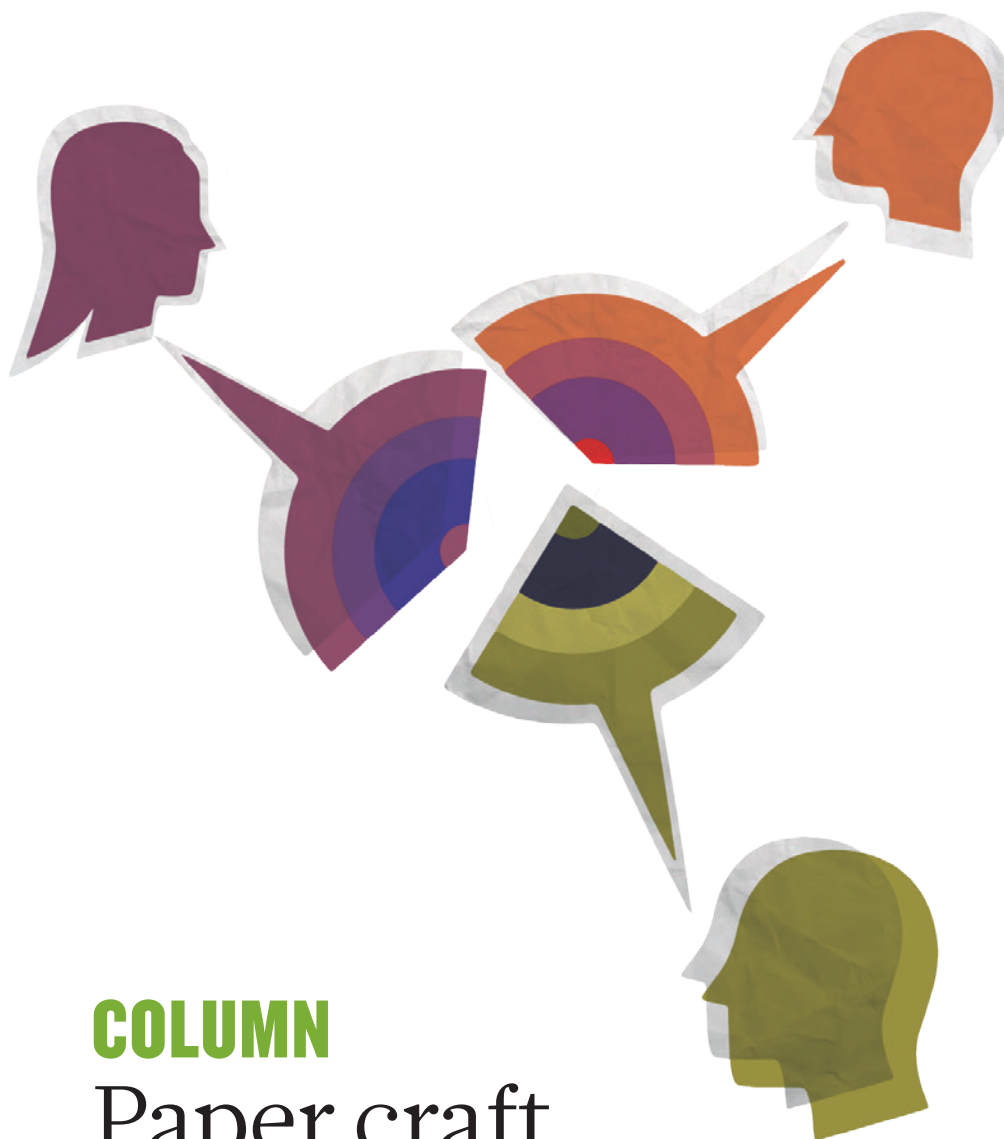
1. Au, K. F. *et al. Proc. Natl. Acad. Sci. USA* **110**, E4821–E4830 (2013).
2. Durruthy-Durruthy, J. *et al. Nature Genet.* **48**, 44–52 (2016).
3. Hangauer, M. J., Vaughn, I. W. & McManus, M. T. *PLoS Genet.* **9**, e1003569 (2013).
4. Cabili, M. N. *et al. Genome Biol.* **16**, 20 (2015).
5. Jinek, M. *et al. Science* **343**, 1247997 (2014).
6. Chu, C. *et al. Cell* **161**, 404–416 (2015).
7. McHugh, C. A. *et al. Nature* **521**, 232–236 (2015).
8. Minajigi, A. *et al. Science* **349**, aab2276 (2015).
9. Novikova, I. V., Hennelly, S. P. & Sanbonmatsu, K. Y. *Nucleic Acids Res.* **40**, 5034–5051 (2012).
10. Spitale, R. C. *et al. Nature Chem. Biol.* **9**, 18–20 (2013).

CAREERS

INTERDISCIPLINARITY Writing together creates clear goals **p.429**

CONSULTING Analysts think like scientists, but have shorter deadlines **p.429**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



is important to write down and follow ground rules on how the author list will be determined, the drafts composed and the editing and revision process carried out. It is especially important to maintain contact among collaborators and ensure that all co-authors are on board with changes to the manuscript. More generally, it is helpful to remember that ultimately your research paper should tell a compelling story — and that this story is the principal, tangible result of the group's work.

ANDY BAKER/GETTY

MAKE THE 'WRITE' DECISIONS

Once collaborators have agreed on ground rules and authorship questions (see 'First things first'), they immediately need to agree on a 'composing author' — the team member who is most familiar with the project details and most likely to have the deepest and broadest perspective. This teammate maintains the master copy of the manuscript and incorporates input from each co-author.

It is most effective if the same person serves as composing author from the project's launch to publication. There are usually many rounds of editing at each stage of the manuscript's life, and often the same points arise at different stages; if one person is shepherding the manuscript throughout, it is easy for them to recall the reasons behind various editing decisions if similar points are raised again. Having a single composing author also helps to establish a coherent, consistent voice.

Of course, life is complicated, and it may not be possible for one person to oversee the whole project. If the original composing author must step down, the team should explicitly agree on a replacement. The manuscript should be 'owned' solely by that person going forward.

We always urge our lab groups and collaborators to start writing before the research project is complete. A paper usually begins with an explanation of the motivation for the project and a survey of earlier related work. What better time to create this section than before you have fully plunged into the study?

Also, doing this bit at an earlier stage helps to delineate the paper's logical path and to provide a guide for the research ahead. When the survey of earlier work is prepared in advance, it is easier to foresee, for instance, what might need to be measured to choose between possible alternative interpretations of the data.

It is also useful to create an overall outline for the paper early on. The composing author should write it, discuss it with the team and reach consensus. ►

COLUMN

Paper craft

Consensus with co-authors is vital when writing up research, say **Dmitry Budker** and **Derek Jackson Kimball**.

Most scientific manuscripts today have anywhere from two to thousands of co-authors. Large collaborations — generally defined as more than a couple of dozen members — usually establish formal structures to govern manuscript writing, but smaller groups often develop their writing rules in an ad hoc way. As a result, we have seen — and participated in — countless small collaborations that struggle with such

fundamental issues as editing their manuscript and deciding on the author list. Here we offer some general rules for small groups on how best to craft a scientific manuscript.

Writing a research paper is not necessarily formulaic, but we have learned some useful principles and guidelines from our own and others' experiences that will help to avoid many such obstacles and produce clear, engaging and readable text. From a practical standpoint, it

► Such a framework helps to crystallize the goals of the project, forces the group to address key scientific questions from the outset and keeps the work focused. More than once, we have completed a study and reconfigured the experimental apparatus for some new investigation, only to discover while writing the paper that we had overlooked a detail that requires more data. The outline can help to avert this misstep.

In some cases, it can work well if the first drafts of different manuscript sections are composed by different team members in accordance with their strengths — a theorist could write a section on calculations and an experimentalist could write a section on measurements, for example.

If your collaboration uses this model, the composing author should be the one to incorporate the different sections into a coherent story.

COMMUNICATION AND CONSENSUS

The decision on whether to add, delete or change existing content in the master copy calls for constant communication between the composing author and each co-author. As the composing author revises the master copy, he or she must send each new version to every co-author along with explanations of the changes. No one except the composing author should edit the master copy — co-authors need to send suggestions, corrections and extra pieces of text to that person.

By the same token, the composing author should acknowledge — although not necessarily incorporate — all such input from co-authors and, ideally, launch a discussion to reach agreement. From the co-author's point of view, it can be extremely frustrating to read an updated draft of the master manuscript and see the very problem that he or she had pointed out earlier — we have seen this happen often. When a co-author asks why the problem was not addressed, the composing author usually says only that he or she disagreed with the suggestion or did not find a good way to change the text. These are the very reasons to have a discussion.

The entire collaboration must agree on a first draft and on all subsequent drafts. We have both co-written papers that went through scores of serious revisions, a process that required a lot of patient, careful work over an extended period. In some of our collaborations, composing authors solicited and incorporated some edits from co-authors without sharing the final version of the manuscript with the entire team.

Co-authors have the right to expect and demand that the composing author follow the established ground rules during a collaboration. If they discover violations of those rules only after submission or publication, their response may well be to sever ties with the collaboration, which would be a calamitous and entirely avoidable result.

FIRST THINGS FIRST

The thorny problem of the author list

Who should be a co-author of the paper?

Two key questions we ask are: 'did the researcher make a meaningful contribution to the project?' and 'does the researcher understand the complete work well enough to explain and defend it to colleagues?'. Becoming a co-author of a paper is a serious commitment: that person's scientific reputation is now wedded to the validity of the paper. We have found that it makes sense to be reasonably inclusive: generally, people who feel that they do not meet the threshold for co-authorship will excuse themselves from the author list. It is much more damaging to the collaboration to leave someone off the author list who feels that they should be included than it is to include someone who has made marginal contributions. The omission can create hard feelings and possibly violate scientific ethics.

The order of authors on a collaborative scientific publication is often a source of

contention within the group, especially because different scientific communities have different and often unwritten rules. The high-energy-physics community, for example, tends to adhere to a strict alphabetical order of authors, whereas in atomic physics, students and postdocs who are central to the project are usually listed first and the principal investigator appears last.

Think about the author order from the start. What do you do if there is more than one student or a student and a postdoc who contributed equally? And what do you do if the work is a collaboration among several laboratories? Which laboratory goes first? Should the authors be grouped by laboratory, or should there be some other criterion?

Explicit descriptions of each author's role in the work can make author-order discussions less fraught. 'Author contribution' sections should be included even for journals that do not require them. These should spell out who did what on the project and who is responsible for which parts of the manuscript. When a collaboration cannot reach consensus about author order, the principal investigator should step in to make those decisions and provide detailed reasoning for the verdict. If there are multiple principal investigators, they must work out the order between them.

Ultimately, co-authors should not worry about this too much. We have witnessed heated arguments over who should be the first author; but 10 or 20 years later, we can see that it did not really matter that much in the end. **D.B. & D.J.K.**



In any case, if you are named composing author, you must consistently keep co-authors in the loop. Excluding them creates a lot of hard feelings and may also violate the norms of scientific ethics.

Everyone on the team needs to be available and responsive. Work comes to a halt when co-authors disappear into a 'black hole' in which e-mails and phone calls go unanswered.

We recommend defining a maximum response time — generally a few days — during which co-authors must at least acknowledge receipt of a communication. We also recommend that the composing author maintain consistent availability for the entire course of the writing project, which in some cases can be a year or longer.

Whether there are 2 authors or 22, each co-author should read the entire final manuscript and explicitly agree that the paper is

ready for journal submission or for posting to an e-print archive. The same holds true when a paper has been submitted and accepted, and is undergoing final edits — the full team needs to read and agree on the final version before publication. This is not only a courtesy to authors, it is also a cornerstone of scientific ethics, not to mention journal-policy and legal considerations.

We emphasize that these guidelines are not arbitrary: they originate from stressful and frustrating experiences that we have both lived through and witnessed. ■

Dmitry Budker is a professor of physics at the Johannes Gutenberg University in Mainz, Germany, and the University of California, Berkeley, and **Derek Jackson Kimball** is a professor of physics at California State University — East Bay in Hayward.

COLUMN

Tracked changes

RICHARD NOVAK

Writing together bridges disciplines, says **Eli Lazarus**.

Not long ago, I co-wrote a grant with a small army of collaborators. As our deadline approached, we were circulating several e-mails a day among 24 people at 8 different institutions around the United Kingdom. Each 'reply-all' message laboured under the size of the latest attachment. Time stamps reflected ever-blearier hours. Our ambition for a major interdisciplinary consortium was at stake.

Our project called for researchers from the natural sciences (in hydrology, soil, ecology, sediment transport, water chemistry and gas exchange) to join colleagues from the social sciences (in economics, politics, policy and governance) to study changes in environmental resources at a national scale. We would collaborate to examine managed landscapes holistically, as integrated systems of natural processes and human activities. We would be a supergroup united by data sets, computer models, statistical analyses and case studies.

As iterations of the proposal rolled through our inboxes, it occurred to me that for all our use of future tense and conditional clauses, the promise of what we would do if funded, this — the collaborative act of writing the proposal — was really where our interdisciplinary work was happening. The requirement of writing a joint proposal was forcing us to transcend the shorthand of our respective disciplines.

The proof was in the rainbow of accumulated tracked changes and marginalia. "We're going to drive this model with results from that one, right?" someone asked in electric green. Yes. Terms familiar to one contributor were foreign to another. Clauses such as "construct a platform for linked subroutines capable of accounting for fluxes through the selected grid space" disappeared. "Deliberately abstracted" became "simplified", and "dynamically integrated comprehensive framework" became "system".

Co-authors haggled over content. Does a group of integrated models comprise a single model? Are we simulating or predicting or forecasting? Should we emphasize the elements that make the work applied, or the unresolved dynamics that make it exploratory? Each description, each choice of evidence, each reference was weighed. Debates about what stayed in a draft and what did not were essential to our internal process of deciding what we actually wanted to do. They propelled simpler, clearer writing — to everyone's benefit. By the time our submission was

ready, what had started as a sprawling pitch had transformed into a compelling plan. The steps, however technically complicated, felt like items that we could tick off of a to-do list.

This grant proposal marks my fourteenth formal interdisciplinary undertaking. Each one has been an education. You hear the language of another discipline: its phrases and idioms; its favourite verbs and adjectives. You learn how colleagues from other backgrounds construct and solve problems. You gain a sense of the ideas that interest them, and an appreciation of why. You hear the language of your own discipline as if it were someone else's. You revisit your own preferences for puzzles and questions.

You strike your foregone conclusions, hang up your implicit assumptions and begin to build — and rebuild — explanations and arguments from scratch.

Every day, I read posts from institutions around the world soliciting ways of cultivating interdisciplinary initiatives. But I have found that strategic discussions about interdisciplinary research rarely come down to the practicalities of the research process.

Interdisciplinary research is too often described in terms of lofty abstraction, a collaborative effort that sounds more magical than methodical. Writing together makes projects real. It requires negotiations between disciplines and individuals (see *Careers Column* page 427). The exercise closes conceptual distances that conversations leave wide open.

The next time I am invited to an interdisciplinary workshop to galvanize links between researchers from disparate specialties, I might ask the organizers how much collaborative writing is on the agenda. If the event is mine to organize, I'll scrap hours of showcase presentations. Attendees will write. I'll offer a handful of relevant prompts: calls for research proposals from a funding agency, advertisements for an interdisciplinary conference, the announcement of a journal's special issue. Participants will establish ground rules for a safe creative space — and we will enjoy the alchemy of what emerges. ■

Eli Lazarus is a lecturer at Cardiff University, UK.

TRADE TALK
Science strategist

Gautham Venugopalan, an analyst at the consulting firm Gryphon Scientific in Takoma Park, Maryland, describes switching from a post as a bioengineer who writes scientific papers to one as a consultant who assesses the literature to inform policy decisions.

How did you learn of jobs outside academia?

I never intended to go into an academic career. As a graduate student researching cancer biophysics, one of the reasons I chose my lab at the University of California, Berkeley, was because people there had gone on to do a variety of things: management consulting, freelance journalism, starting a company. Also, my adviser had an interest beyond the lab: he had done a White House fellowship. He would come back and say, "These are the scientific issues, and here are these societal components, political components and budget components".

I started a non-profit with some friends and then did a policy fellowship at the US Department of State. At a networking event, I met someone from Gryphon and ended up interviewing for a job there.

What do you do now?

Gryphon is a small research and consulting firm with 50 or 60 people. We use scientific analysis to advise people on national security and global health issues.

How do you apply your training?

One thing you hear a lot when you leave academia is, "Don't you wish you could use your scientific training more?" And right now I do use it. There is still a lot of uncertainty, but instead of spending five years finding out a piece of that, you spend five months generating a report, which is a different challenge but uses a lot of the same skills.

You use evidence to justify your conclusions, and sometimes you use scientific principles to test your assumptions or to provide a range of estimates. You need to know how publishing works and how evidence works if you want to make evidence-based policy. ■

INTERVIEW BY MONYA BAKER

This interview has been edited for length and clarity. See go.nature.com/ifbowr for more.

THERE IS A BEEP

One-way traffic.

BY FILIP WILTGREN

There is a beep.

An applied neurophysicist walks into a bar. It's a nice bar, red leather upholstery, jukebox in the corner, bartender in a white dinner jacket. More of a lounge, really. And this neurophysicist walks into the place like he owns it.

Maybe he does.

"Joe," he says, sitting down, "fix me up a stiff one."

And the bartender bends behind the bar and comes up with one of these silver shakers that rattle like a million dollars on ice and he rattles it and breaks it and there's this rich brown, velvety fluid emerging from the shaker and sloshing in slow motion into a heavy, cut-crystal glass.

The physicist goes to the lavatory. You know it's the lavatory because it says lavatory right on the door. Not toilet, not WC. Lavatory.

That's because the physicist is British and hasn't taken to the language of wherever it is he is.

So now the neurophysicist is gone. The lavatory door swings shut behind him.

There is a beep.

A soon-to-be-Nobel-laureate neurophysicist walks into a bar. It's a nice bar, the kind with red leather upholstery with big round buttons. Sinatra's crooning from a jukebox hooked into a Bang & Olufsen sound system. The bartender looks classy, like he makes more in an hour than most chief executives.

"Joe," says the physicist, "fix me a stiff one."

And the bartender does his magic, and the liquid flows and sloshes, and there's a smell like old oak and cinnamon. Beads of dew condense on the glass and flow down to the physicist's hand. He lifts the glass to his mouth, tilts it. It tastes like victory. The neurophysicist smacks his lips in appreciation. He raises his finger to order another but then decides against it. He pulls out a billfold and removes an Irish €10 note. James Joyce winks at him from the paper. That feels wrong, somehow.

The physicist walks to the door, but it's locked. He gets a slightly worried look on

his face and turns to the lavatory instead. The lavatory door swings shut.

There is another beep.

A somewhat confounded neurophysicist walks into a bar. It's a nice bar, retro styled with lacquered benches the colour of maraschino cherries. There's a mahogany bar, a balding bartender, a jukebox playing 'Auld Lang Syne' in the background.

"Hey," says the physicist, "fix me one of those."

He points to a row of bottles standing at attention on a shelf behind the bar. They're all Bushmills 21-year-old single malt. Every second one is half empty.

The bartender bends down and hauls out a new bottle of malt. The seal cracks like fireworks. The liquid sloshes in slow motion into a glass full of ice cubes, condensation forming on the outside. A single drop runs down the side of the glass, amplifying the perfect, lip-smacking clarity of the liquid behind it. The physicist picks up the glass.

It feels completely dry, like paper.

ON NATURE.COM
Follow Futures:
@NatureFutures
f go.nature.com/mtoodm

The physicist tilts the glass. The whiskey stays the same, frozen in a half swirl, the ice cubes almost

clinking. It looks like an ad in an upscale lifestyle magazine.

"Bugger, bleeding, bugger, hell," the physicist says and throws the glass. It shatters the wall-length mirror behind the bar. The bar's door won't open, the lavatory door swings shut.

There is a rather cut-off beep.

An anxious neurophysicist walks into a bar. It's a nice bar, the kind where you'd take your date if you had the right salary. There's a bartender, a jukebox, a tinted, wall-length mirror. There's a lavatory.

The physicist turns and looks at the door he came through. There's a small, red, corroded sign near the handle. It reads: "One way only".

"Hey," says the physicist, "you've got another door in this place?"

The bartender raises his arm and points at the lavatory.

"Crap," says the physicist.

There is no beep.

There's this bar, and there's a physicist in it, and the bar's got two doors, one that opens and one that don't. The one that opens is marked "Crapper".

The physicist is confused. He doesn't need to go to the crapper, he needs a drink and he doesn't know why he needs it.

"Hey," he says to the bartender, "fix me up a shot, will ya?"

The bartender bends down behind a row of antique silver drink shakers and pulls out a shotgun. It goes well with the tweed cap he's got on. He points it at the physicist.

"Crap," he says and pulls the trigger.

The leather upholstery is decorated in splattered-neurophysicist chic.

There is no beep, there is no bar, there is no neurophysicist. There is only a failed experiment, a final ECG tone, a long, drawn-out, unpleasant investigation, and a case study on the importance of following established safety protocols. ■

Filip Wiltgren is a communications officer at Linköping University by day and a writer, game designer and worrier by night. He lives in Sweden with his wife and daughters.

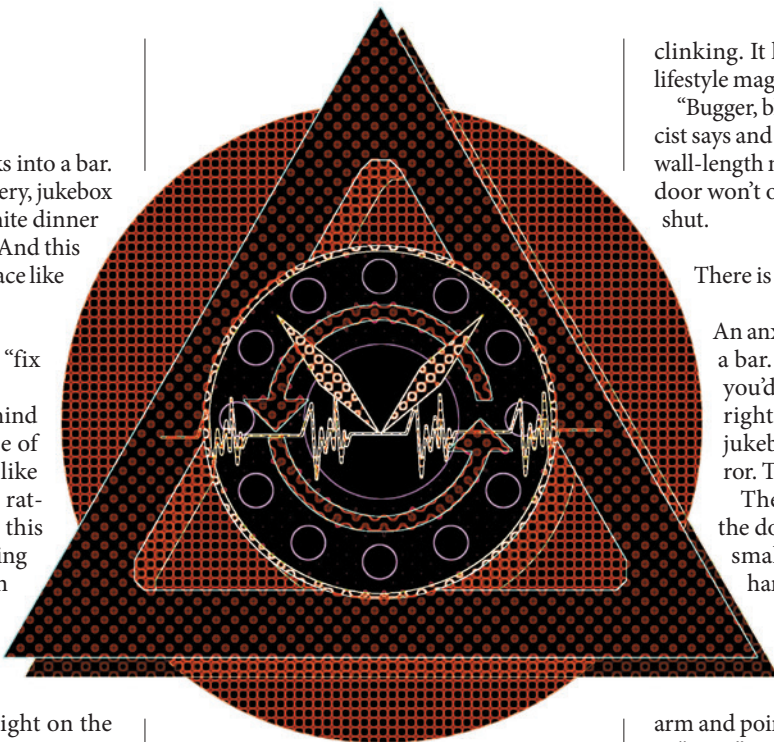


ILLUSTRATION BY JACEY